2016

Molecular evolution of opsins, a gene responsible for sensing light, in scallops (Bivalvia: Pectinidae)

Anita J. Porath-Krause
Iowa State University

Molecular evolution of opsins, a gene responsible for sensing light, in scallops (Bivalvia: Pectinidae)

by

Anita J. Porath-Krause

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Ecology and Evolutionary Biology

Program of Study Committee:
Jeanne M. Serb, Major Professor
Maura A. McGrail
Dean C. Adams
Anne M. Bronikowski
Stephan Q. Schneider

Iowa State University

Ames, Iowa

2016

DEDICATION
For my son

TABLE OF CONTENTS

TABLE OF CONTENTS iii
LIST OF FIGURES v
LIST OF TABLES vi
ACKNOWLEDGEMENTS vii
ABSTRACT viii
CHAPTER I. GENERAL INTRODUCTION 1
Overview 1
Gene duplications and their fate 1
Opsin 3
Visually-mediated behaviors and scallops 5
Scallop eyes 8
Dissertation outline 10
References 11
CHAPTER II. STRUCTURAL DIFFERENCES AND DIFFERENTIAL EXPRESSION AMONG RHABDOMERIC OPSINS REVEAL FUNCTIONAL CHANGE AFTER GENE DUPLICATION IN THE BAY SCALLOP, *ARGOPECTEN IRRADIANS* (PECTINIDAE) 20
Abstract 21
Background 22
Methods 26
Results 33
Discussion 39
Conclusions 44
Acknowledgments 46
References 46
CHAPTER III. DIFFERENT EVOLUTIONARY TRAJECTORIES BETWEEN GQ-OPSIN PARALOGS SAMPLED ACROSS SCALLOPS 77
Abstract 77
Introduction 78
Methods 82
Results 95
Discussion 104
References 109

CHAPTER IV. THE DISADVANTAGE OF IGNORING ALLELIC VARIATION WITH AN EMPHASIS ON SCALLOP PHOTOSENSORY ALLELES 137
Abstract 137
Introduction 138
Methods 144
Results 147
Discussion 149
References 153
CHAPTER V. CONCLUSIONS 166

LIST OF FIGURES

Figure 1-1 Topographical model of G_q-opsin 18

Figure 1-2 A sagittal cross-section of the scallop eye showing all ocular components present at the adult stage 19

Figure 2-1 Amino acid alignment of G_q-opsins from scallop (Air-OPNGq1-OPNGq4) and squid, *Todarodes pacificus* (Tpa-OPNGq1) 66

Figure 2-2 Maximum likelihood (ML) topology of G_q-opsins 67

Figure 2-3 38 amino acid sites predicted to interact with chromophore in G_q-opsins 69

Figure 2-4 Expression profiles of scallop G_q-opsin genes across three tissues from a single light-treated animal 70

Figure 2-5 An evolutionary hypothesis describing G_q-opsin duplications in Pectinioidea ... 71

Figure 2-S1 Bayesian inference phylogram of G_q-opsins 73

Figure 2-S2 Maximum likelihood phylogram of G_q-opsins 75

Figure 2-S3 Fifty base pair alignment of 5'- and 3'-UTRs from the four scallop G_q-opsins .. 76

Figure 3-1 G_q-opsin gene tree 128

Figure 3-2 Sequence variation of G_q-opsins 129

Figure 3-S1 CODEML tests 130

Figure 4-1 OPNGq1 and OPNGq2 amino acid sequence variation across scallops 165

LIST OF TABLES

Table 2-1 Percent similarity (below diagonal) and RMSD (above diagonal) of scallop (Air) and squid (Tpa) proteins 57

Table 2-2 Sequence and structural motifs in scallop (Air) and squid (Tpa) G_q-opsins..... 58

Table 2-S1 Primers used to amplify scallop G_q-opsins and intergenic region between *Air-opnGq3* and *Air-opnGq4* 59

Table 2-S2 G_q-opsin sequences included in the phylogenetic analysis..... 60

Table 2-S3 Ramachandran plot values and C-scores for top G_q-opsin models..... 64

Table 3-1 Primers used to amplify scallop G_q-opsin paralogs 116

Table 3-2 Species included in analyses..... 117

Table 3-3 Rates of evolution for each test..... 118

Table 3-4 Testing for variable rates of evolution between *opnGq1* and *opnGq2* 119

Table 3-5 Putative spectral tuning sites..... 120

Table 3-6 Positions of interest..... 121

Table 3-S1 Duplicate, Depth, Behavior information..... 122

Table 3-S2 Codon-based test of diversifying selection 123

Table 3-S3 Codon-based test of purifying selection 124

Table 3-S4 Branch-site method of site-specific selection 125

Table 3-S5 Relationship between behavior and sequence variant 126

Table 3-S6 Relationship between depth and sequence variant..... 127

Table 4-1 Species, number of different nucleotide and amino acid sequences, and GenBank accession numbers of nucleotide sequences 160

Table 4-2 Single nucleotide polymorphisms and percent pairwise identity of nucleotide sequences 163

Table 4-3 dN/dS ratios from randomized samplings of available nucleotide sequences... 164

ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee, Drs. Dean Adams, Anne Bronikowski, Maura McGrail, and Stephan Schneider for their guidance and support. I would especially like to thank my committee chair, Dr. Jeanne Serb, for her patience and mentorship. To my graduate cohort and good friends, who were more support than they could ever imagine. And finally, I would like to thank my family, especially my extremely supportive husband, Bob.

ABSTRACT

Genetic diversity can cause drastic effects on phenotypes and is commonly the result of a gene duplication event. Gene duplication and subsequent functional divergence of opsins, a G-protein coupled-receptor (GPCR), have played an important role in expanding photoreceptive capabilities of organisms by altering what wavelengths of light are absorbed by photoreceptors (spectral tuning). Relatively few studies have been devoted to exploring the role of opsin duplication and evolution in non-arthropod invertebrates, and even fewer have integrated all the potential genetic diversity of opsins. In this dissertation, I utilized the photosensory system of the scallop, a marine bivalve, to study the evolution and expansion of the genetically diverse opsins, and demonstrate the complicated nature of G_q-opsin diversification after gene duplication. First, I explored how opsin paralogs diversify in function and evolutionary fate by characterizing four rhabdomeric (G_q-protein coupled) opsins in the scallop, *Argopecten irradians*. Using a phylogenetic framework, I showed a pattern consistent with two rounds of duplication generating the four paralogous G_q-opsins in scallops. Differential expression of the four G_q-opsins across ocular and extra-ocular photosensitive tissues suggested that the G_q-opsins are used in different biological contexts in scallops, while protein modeling reveals variation in the amino acid composition, suggesting the four G_q-opsin paralogs may absorb different wavelengths of light. Second, I investigated how two G_q-opsin paralogs differentiate after a duplication event across the scallop family, Pectinidae. By comparing the rates of evolution between paralogous clades, I demonstrated both paralogs are under purifying selection, yet maintained at rates that are significantly different. I showed that one amino acid position, which is not considered a putative spectral tuning site, stands out as a strong candidate to explain the source of selection driving the difference in evolutionary rates. Finally, I discussed the current role of allelic variation in sensory systems and described how alleles are often discarded in

studies of molecular evolution. I demonstrated the breadth of possible allelic variation within an individual and stressed the potential of cryptic genetic variation in the evolution of organisms by examining the allelic variation in G_q-opsins sampled across 34 bivalve species.

CHAPTER I

GENERAL INTRODUCTION

Overview

Genetic diversity has profound effects on the survival and evolution of organisms as it is critical for the ability to cope with constantly changing environments, adapt to local environments, and increase resistance to extinction in natural populations (Webster *et al.*, 1972; Yang & Patton, 1981; Barrett & Schluter, 2008). Proteins that are required for vision are an effective system to examine how genetic diversity, driven by environmental factors, such as light availability, influences the evolution of an organism. In animals with eyes capable of spatial vision, combinations of visual pigments, which are comprised of a G-protein coupled-receptor (GPCR), called opsin, and a vitamin A-derived chromophore, can expand the visual capabilities of organisms by altering what wavelengths of light are preferentially absorbed by photoreceptors, which can have profound effects on the evolution of organisms (Webster *et al.*, 1972; Yang & Patton, 1981; Yokoyama, 2002; Barrett & Schluter, 2008; Hofmann & Carleton, 2009). Amino acid changes in visual pigments can result in changes in sensitivity to specific wavelengths of light (e.g., Sugawara *et al.*, 2005). This is quantifiably measured a **wavelength absorption,** the wavelength at which the absorbance of light is the greatest (λ_{max}) (Yokoyama *et al.*, 2008).

Gene duplications and their fate

Selection acts on the genetic variation underlying different phenotypes. Most new gene content and genetic variation are not from a *de novo* origin (Jacob, 1977 but see Cai *et al.*, 2008;

Knowles & McLysaght, 2009), but are from the process of gene duplication (Ohno, 1970). Gene duplicates can result from several events including: whole-genome duplication events (Tank *et al.*, 2015; Vallejo-Marin *et al.*, 2015), segmental duplications (Cheung *et al.*, 2003), retrotransposition (Xiao *et al.*, 2008), unequal crossing over (Ohno, 1970), and most commonly, duplicative transposition (Lee & Van der Ploeg, 1987; McCulloch *et al.*, 1997; Freeling, 2009). Gene duplicates have the potential to supply additional substrate for evolution through the retention of the original function but also allow for the evolution of a new function (Ohno, 1970). The possibility of obtaining and retaining a new function is based on the fate of the duplicate gene, which can be grouped into four categories: gene conservation, pseudogenization, subfunctionalization, and neofunctionalization.

Gene conservation increases the number of genes coding for a protein by retaining both duplicate genes, and both loci maintain the original function (Zhang, 2003). Retention of the original function in both gene copies can result in an increase the amount or dosage of expression of the genes. Studies have demonstrated increasing the amount of expression can be beneficial by, for example, increasing tumor suppressors (Matheu *et al.*, 2004) and reducing disease susceptibility (Gonzalez *et al.*, 2005). However, increasing the dosage of gene products has the potential to be harmful and even lethal to the organism (Qian *et al.*, 2010). This result is much more difficult to demonstrate because those individuals with harmful or lethal dosage effects would quickly be selected out of the population or in ideal situations, experience a reduction in dosage of gene products.

Pseudogenization is the most probable fate of gene duplication events. Pseudogenization results in non-functional genes that accumulate in the genome and become junk DNA. Relatively “young” duplicated genes that were pseudogenized can be aligned to the original

sequence to give insight into historical patterns of that gene or gene family (Cortesi *et al.*, 2015), and on rare occasions, be revived into functional genes (discussed in Ohno, 1970; Zhang, 2003).

Subfunctionalization is the division of ancestral function among duplicated loci. In the strictest sense, subfunctionalization results in the division of expression domains, usually coupled with the loss of partial function from one or both duplicates so that the ancestral function is divided between the two duplicates (Lynch & Force, 2000; Tank *et al.*, 2015; Vallejo-Marin *et al.*, 2015). Some use the term subfunctionalization to describe the partition of gene function as an escape from adaptive conflict (Piatigorsky & Wistow, 1991; Cheung *et al.*, 2003) or a separation in temporal or spatial expression (Force *et al.*, 1999; Spady *et al.*, 2006; Xiao *et al.*, 2008).

Neofunctionalization is a pair of duplicate genes where one possesses a new, selectively beneficial function that was absent in the population before the duplication (Ohno, 1970). Neofunctionalization has also been described as mutations that accumulate at the redundant locus due to drift (Dykhuizen & Hartl, 1980; Lee & Van der Ploeg, 1987; McCulloch *et al.*, 1997; Freeling, 2009). The new version of the duplicated gene can be advantageous to the organism.

Opsin

Opsins encode a class of G-protein coupled receptors (GPCRs) that span the membrane of photoreceptor cells (Fig. 1-1). When an opsin protein is covalently bound to a light-sensitive chromophore, the molecule acts as a photopigment that is sensitive to a specific portion of the light spectrum. When the chromophore absorbs a photon light, it undergoes photoisomerization, changing its conformation from an 11-*cis* to an all-*trans* state. This conformational change then modifies the conformation of the opsin, initiating the phototransduction cascade, which results in

either a depolarization or hyperpolarization of the membrane potential in the cell (Smith *et al.*, 1991; Zhang, 2003). Under a phylogenetic framework, there are four main opsin classes estimated to be present in the last common bilaterian ancestor: c-opsins, tetraopsins, G_q-opsins, and the most recently described, xenopsin (Ramirez *et al.* 2016). Opsin classes largely correspond to their association with specific heterotrimeric G proteins that contains an alpha, beta, and gamma subunit (Feuda *et al.*, 2014; reviewed in Porter *et al.*, 2012). More recently however, opsin classes are commonly defined by their phylogenetic relationship and may associate with more than one G-protein type (e.g., the c-opsin class contains opsins that associate with both G_{s-} and $G_{i/o}$ pathways) (Rameriz *et al*. 2016). Opsins from each of the four classes can be expressed-in more than one tissue type and can be used in multiple biological contexts including vision (Speiser & Johnsen, 2008), regulating circadian rhythm and photoperiod (Hattar, 2002; Arendt *et al.*, 2004), shadow detection (Kennedy, 1960; Lesser *et al.*, 2011), and melanophore detection (Provencio *et al.*, 1998).

Opsins have been identified in every major animal phyla (Arendt & Wittbrodt, 2001; Arendt *et al.*, 2004; Purschke *et al.*, 2006; Kozmik *et al.*, 2008; Battelle *et al.*, 2015; but see Mohamed, 2003; Mohamed *et al.*, 2007) and phylogenetic placement suggests multiple opsin paralogs were present in the last common bilaterian ancestor, suggesting opsins have long played a role in photosensitivity (Ramirez *et al.* 2016). Vertebrate opsin evolution has been well-described in the context of gene duplication (color vision), convergence, and adaptation (Blackshaw & Snyder, 1997; Blackshaw & Snyder, 1999; Provencio *et al.*, 1998; Dulai *et al.*, 1999; Yokoyama, 2000; Wang *et al.*, 2004; Terai *et al.*, 2006; Hunt *et al.*, 2009), through the exploration of signatures of selection within the opsin sequence (Meredith *et al.*, 2013), by comparing multiple opsins across more than one species (Spady, 2005; Smith & Carleton, 2010);

and by making large-scale comparisons to show adaptation to new environments or habitats (Yokoyama *et al.*, 2008). Studies of arthropod opsin evolution also appear to have increased in interest (Carulli & Hartl, 1992; Briscoe, 2001; Spaethe, 2004; Porter *et al.*, 2006; Battelle *et al.*, 2015); however, the molecular evolution of opsins in molluscs, which are the second most speciose animal phylum, are much less characterized. By incorporating studies of opsin molecular evolution across an entire family, or even across an entire order of molluscs, broad-scale patterns may contribute to critical insight regarding the evolution of opsins before and after the bilaterian split, and may demonstrate how the evolution of molluscan opsins may differ from vertebrates and even other invertebrate lineages.

Visually-mediated behaviors and scallops

Vision is an important component for the survival of many animals. Eyed organisms rely on visual cues from the environment to regulate many behaviors including foraging (Dulai *et al.*, 1999; Gumbert, 2000), recognizing predators (Barcellos *et al.*, 2007), avoiding seemingly unpalatable prey (Kraemer *et al.*, 2015), navigation during migration (Wiltschko & Wiltschko, 1991), recognizing conspecifics (Baracchi *et al.*, 2013), and locating habitats (Hamilton & Koch, 1996). Visually-mediated behaviors are often coupled with- or complemented by other sensory systems (Wiltschko & Wiltschko, 1991; Wang *et al.*, 2004; Chittka & Raine, 2006; McCormick & Manassa, 2007; van Breugel *et al.*, 2015) to better sample the environment. Identifying evolutionary mechanisms that refine visually-mediated behaviors is important to understand the adaptation of sensory systems, and how changes to sensory systems play a role in animal evolution (Briscoe & Chittka, 2001; Yokoyama *et al.*, 2008). For example, LWS (a type of opsin used for color vision) gene divergence driven by female preference may have contributed

to male nuptial coloration and subsequent speciation of cichlids found in Lake Victoria (Terai *et al.*, 2006). The adaptation of visual pigments to contain a λ_{max} specific to the environment or a specific purpose are referred to as **spectral tuning** and can occur through multiple mechanisms, such as replacing one type of chromophore with another (e.g., 11-*cis*-retinal vs. 11-*cis*-3, 4-dehydroretinal) or with a nonsynonymous mutation in the opsin protein (reviewed in Yokoyama, 2000). Studies of opsins involved with spectral tuning has been biased toward vertebrate visual systems (Hunt *et al.*, 1998; Spady, 2005; Yokoyama *et al.*, 2008; Sivasundar & Palumbi, 2010), yet due to limitations of expressing opsins *in vitro*, invertebrate opsins, especially those in non-arthropod systems, have been much less characterized. In lieu of direct testing, alternative approaches (e.g., bioinformatic modeling based on crystal structures from closely related organisms, nonsynonymous comparisons between opsins, and expression studies) can be used to estimate which positions may be important for spectral tuning.

In this dissertation, I study scallops, which are an excellent model system to understand what creates and maintains genetic diversity in opsin proteins. These marine bivalves are found in diverse, yet quantifiable photic environments (Minchin, 2003) and exhibit repeated evolution of visually-mediated behaviors, such as gliding, a swimming behavior has independently evolved at least four times in the family (Alejandrino *et al.*, 2011). In addition to being one of the few bivalves that can swim, scallops have the ability to use visual cues to identify preferred habitat (Hamilton & Koch, 1996). While it has been demonstrated that many bivalves are photosensitive, not all photosensitive bivalves have eyes (Kennedy, 1960; Buddenbrock & Moller-Racke, 1953). Scallops are unusual bivalves because they have multiple, single-chambered eyes. Like other eyed animals, scallop eyes express opsin (Speiser *et al.*, 2011), and are also found to express opsins in non-visual structures, including the adductor muscle and the

mantle, which is the soft body wall that protects the scallop's visceral mass (Porath-Krause *et al*., 2016). Both ocular and extra-ocular scallop tissues described here contain not one, but multiple opsins that are the result of multiple duplication events (Serb *et al.*, 2013a; Porath-Krause *et al*., 2016).

Many bivalves are considered immobile or permanently sedentary and do not have eyes; scallops, however, all have eyes and only two of the six behaviors are considered sedentary at the adult stage. The relationship between mobility and the presence of eyes is essential to understanding a visually-mediated behavior. Adult scallops exhibit visually-mediated behaviors which are classified into one of six behavioral groups: byssal attaching, nestling, cementing, gliding, recessing, and free-living (see Alejandrino *et al.*, 2011 for full descriptions). A behavior is categorized based on quantifying the degree or type of movement and attachment (Stanley, 1970), which ranges from permanently sedentary, cementing themselves to rocks or coral (i.e., cementing or nestling), to swimming long distances (i.e., gliding), travelling up to 30 meters (Joll & Caputi, 1995). Byssal-attachers retain the ability to produce a temporary protein fastening, the byssus, and have the ability to release and reorient the attachment if needed. Nestling species also attach with a byssus, but become permanently confined within a cavity of coral, specifically coral from the genus *Porites*, and remain sedentary as the coral grows around the scallop. Cementers will permanently fasten onto hard substrates through the secretion of new shell material. Gliders are very mobile as they are able to swim greater than five meters with a gliding component. They rarely attach to substrate with their byssus as adults. Free-living species passively occupy a position on or partially covered in soft or sandy substrates. These species are also able to swim, but do not use the glide component. Recessers excavate a cavity in the sediment to fully or partially conceal themselves. All non-cementing or permanently sedentary

species are considered mobile as they have the ability to swim. Behavioral evidence demonstrates how the free-living scallop, *Argopecten irradians*, is able to use visual information to orient their swimming towards a preferred location (Hamilton & Koch, 1996). Other evidence shows scallops will close their shells in response to movement without the object casting a shadow on the scallop's eyes (Buddenbrock & Moller-Racke, 1953) suggesting the scallop can see and react to the object.

Scallop eyes

The adult scallop possesses up to 200 image-forming, single-chambered eyes along the edge of the mantle tissue that line the left and right valves of the animal. These eyes contain a cornea, lens, two layers of retina, and a spherical concave argentea mirror that lines the back of the eye which is used to focus the light and reflect it on the proximal and distal retinas (Fig. 1-2) (Land, 1965). The presence of a mirror is only found in a handful of other animals, including the spookfish *Dolichopteryx longipes* or (Wagner *et al.*, 2009), and three classes of crustaceans (Cronin *et al.*, 2014). After metamorphosis in the juvenile stage, scallop eyes are small papillae containing undifferentiated retina that are likely restricted to simple directional photoreception (Audino *et al.*, 2015). As the scallop matures and enters the final stages of juvenile development, the eyes develop into ocular organs lining the middle mantle fold containing differentiated structures, which harbor the ability of spatial vision (Audino *et al.*, 2015). Eye morphology, such as the lens and corner shape, retina depth and width, and distance of tapetum to the proximal retina, varies greatly between adult scallop species (Malkowsky & Götze, 2014; Malkowsky & Jochum, 2014) and may be correlated with habitat (Malkowsky & Götze, 2014) or visually-mediated behavior (Speiser & Johnsen, 2008). For example, in scallops with greater

swimming capacity (i.e. exhibit free-living and gliding behaviors), the lens shape tends to be larger, the space between the proximal retina and mirror tends to be smaller, and the photoreceptors are more tightly packed. In contrast, sedentary and permanently cementing species have smaller lenses with larger spaces between the proximal retina and mirror which may increase the retina inter-receptor angles (a measure of optical resolution) (Speiser & Johnsen, 2008).

Physiological and behavioral studies demonstrate that scallop eyes not only detect light and dark (as a shadow response) (Land, 1966), but are capable of spatial vision (Land, 1965; Hamilton & Koch, 1996; Speiser & Johnsen, 2008). Each retina contains a different photoreceptor cell-type, rhabdomeric photoreceptor cells and ciliary photoreceptor cells in the proximal and distal retinae, respectively. The G_o-alpha subunit protein was shown to co-localize in the distal retina while the G_q-alpha subunit protein was shown to co-localize in the proximal retina, suggesting each retina contains the first of the three heterotrimeric subunits required for the phototransduction cascade of the G_o-opsin and G_q-opsin, respectively (Kojima, 1997). Evidence from physiological data suggest that both retinae are used for spatial vision (Speiser *et al.*, 2011), however comparative morphology speculates the proximal retina performs tasks that are used more in scallop species with the ability to swim (Speiser & Johnsen, 2008). While G_o-opsins are found in distal retina of scallops and hypothesized to be used for spatial vision in scallops, G_q-opsins are more often found to be associated with protostome visual capabilities (Porter *et al.*, 2012; Rameriz *et al*. 2016) and much more research is available regarding relationships the between G_q-opsins and invertebrates (Koyanagi & Terakita, 2008). For example, a duplication of G_q-opsin genes was recently documented in the order Pectinoida which includes three bivalve families (Serb *et al.*, 2013a). Sequence divergence and phylogenetic data

revealed the duplication event occurred in an oyster-Pectinoida ancestor, and the likelihood of retaining both genes is associated with either the presence of eyes and/or the degree of mobility (Serb *et al.*, 2013b).

The uncanny genetic diversity, as will be described in detail, found in scallops may reflect the complex nature of molecular evolution of opsins. Scallops' predictable photic environments, visually-mediated behaviors, presence of double-retina eyes, and history of opsin duplication events make this marine bivalve a great system to look for patterns of convergence and divergence, and changes in evolutionary rates that may provide further insight to the process of molecular evolution of scallop opsins.

Dissertation outline

Gene duplication events are incredibly important for natural selection and evolution as they provide a large portion of genetic variation. In this dissertation, I explore G_q-protein coupled opsin gene (G_q-opsin) duplicates, which appear to contribute a fountain of genetic variation for a family of marine bivalves, the scallop (pectinidae). In Chapter II, I feature the discovery and characterization of four G_q-opsin paralogs in one scallop species, *Argopecten irradians*. By comparing the biochemical properties of amino acid residues interacting with the chromophore, the expression levels of each gene, and the spatial expression patterns among light-sensitive tissues in the adult organisms, I demonstrate how neofunctionalization of the G_q-opsin gene copies produced a change in the nearly ubiquitous shadow response in molluscs to a narrowed functional specificity for visual processes in the eyed scallop. For Chapter III, I compare the evolutionary trajectories of two G_q-protein coupled opsin gene paralogs (herein *opnGq1* and *opnGq2* for the gene or the coding region, and OPNGq1 and OPNGq2 for the

protein), *opnGq1* and *opnGq2*, among 34 Pectinoida species of scallop and three closely related marine bivalve families and test if they are under different types of selection. To identify the source of the selection, I test for a correlation between G_q-opsin sequence and two characters that may be important for adaptation: behavior class and photic environment. In the fourth chapter, I stress the importance of cryptic variation in the evolution of organisms, and then demonstrate the breadth of possible allelic variation of *opnGq1* and *opnGq2* nucleotide sequences across the same 34 bivalve species used in Chapter III. Finally, I discuss in Chapter V why the scallop visual and photosensory system has nearly limitless potential to study molecular adaptation and evolution.

References

Alejandrino, A., Puslednik, L. & Serb, J.M. 2011. Convergent and parallel evolution in life habit of the scallops (Bivalvia: Pectinidae). *BMC Evol. Biol.* **11**: 164.

Arendt, D. & Wittbrodt, J. 2001. Reconstructing the eyes of Urbilateria. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **356**: 1545–1563.

Arendt, D., Tessmar-Raible, K., Snyman, H., Dorresteijn, A.W. & Wittbrodt, J. 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science* **306**: 869–871.

Audino, J.A., Marian, J.E.A.R., Wanninger, A. & Lopes, S.G.B.C. 2015. Development of the pallial eye in Nodipecten nodosus (Mollusca: Bivalvia): insights into early visual performance in scallops. *Zoomorphology* **134**: 403–415.

Baracchi, D., Petrocelli, I., Cusseau, G. & Pizzocaro, L. 2013. Facial markings in the hover wasps: quality signals and familiar recognition cues in two species of Stenogastrinae. *Animal Behaviour*. **85**: 203-212.

Barcellos, L.J.G., Ritter, F., Kreutz, L.C., Quevedo, R.M., da Silva, L.B., Bedin, A.C., *et al.* 2007. Whole-body cortisol increases after direct and visual contact with a predator in zebrafish, Danio rerio. *Aquaculture* **272**: 774–778.

Barrett, R.D.H. & Schluter, D. 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol. (Amst.)* **23**: 38–44.

Battelle, B.A., Kempler, K.E., Saraf, S.R., Marten, C.E., Dugger, D.R., Speiser, D.I., *et al.* 2015. Opsins in Limulus eyes: characterization of three visible light-sensitive opsins unique to and co-expressed in median eye photoreceptors and a peropsin/RGR that is expressed in all eyes. *J. Exp. Biol.* **218**: 466–479.

Blackshaw, S. & Snyder, S.H. 1999. Encephalopsin: A Novel Mammalian Extraretinal Opsin Discretely Localized in the Brain. *The Journal of neuroscience*. **19**: 3681-3690.

Blackshaw, S. & Snyder, S.H. 1997. Parapinopsin, a Novel Catfish Opsin Localized to the Parapineal Organ, Defines a New Gene Family. *The Journal of neuroscience*. **17**: 8083-8092.

Briscoe, A.D. 2001. Functional Diversification of Lepidopteran Opsins Following Gene Duplication. *Mol. Biol. Evol.* **18**: 2270-2279.

Briscoe, A.D. & Chittka, L. 2001. The evolution of color vision in insects. *Annu. Rev. Entomol.* **46**: 471–510.

Buddenbrock, W.V. and Moller-Racke, I., 1953. Uber den lichtsinn von Pecten. *Pubbl. Staz. Zool. Napoli*. **24**: 217-245.

Cai, J., Zhao, R., Jiang, H. & Wang, W. 2008. De Novo Origination of a New Protein-Coding Gene in Saccharomyces cerevisiae. *Genetics* **179**: 487–496.

Carleton, K.L. & Kocher, T.D. 2001. Cone opsin genes of african cichlid fishes: tuning spectral sensitivity by differential gene expression. *Mol. Biol. Evol.* **18**: 1540–1550.

Carulli, J.P. & Hartl, D.L. 1992. Variable rates of evolution among Drosophila opsin genes. *Genetics* **132**: 193–204.

Cheung, J., Estivill, X., Khaja, R., MacDonald, J.R., Lau, K., Tsui, L.C., *et al.* 2003. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.* **4**: R25.

Chittka, L. & Raine, N.E. 2006. Recognition of flowers by pollinators. *Current opinion in plant biology*. **9**: 428-435.

Cortesi, F., Musilová, Z., Stieb, S.M., Hart, N.S., Siebeck, U.E., Malmstrøm, M., *et al.* 2015. Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes. *Proc. Natl. Acad. Sci. U.S.A.* **112**: 1493–1498.

Cronin, T.W., Johnsen, S., Marshall, N.J. & Warrant, E.J. 2014. *Visual Ecology*. Princeton University Press.

Dulai, K.S., Dornum, von, M., Mollon, J.D. & Hunt, D.M. 1999. The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Res.* **9**: 629–638.

Dykhuizen, D. & Hartl, D.L. 1980. Selective neutrality of 6PGD allozymes in E. coli and the effects of genetic background. *Genetics* **96**: 801–817.

Feuda, R., Rota-Stabelli, O., Oakley, T.H. & Pisani, D. 2014. The Comb Jelly Opsins and the Origins of Animal Phototransduction. *Genome Biol Evol* **6**: 1964–1971.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.-L. & Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.

Freeling, M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual review of plant biology*. **60**: 433-453.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., *et al.* 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**: 1434–1440.

Gumbert, A. 2000. Color choices by bumble bees (Bombus terrestris): innate preferences and generalization after learning. *Behavioral Ecology and Sociobiology* **48**: 36–43.

Hamilton, P.V. & Koch, K.M. 1996. Orientation toward natural and artificial grassbeds by swimming bay scallops, Argopecten irradians (Lamarck, 1819). *Journal of Experimental Marine Biology and Ecology* **199**: 79–88.

Hattar, S. 2002. Melanopsin-Containing Retinal Ganglion Cells: Architecture, Projections, and Intrinsic Photosensitivity. *Science* **295**: 1065–1070.

Hofmann, C.M. & Carleton, K.L. 2009. Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. *Integrative and Comparative Biology* **49**: 630-643.

Hunt, D.M., Carvalho, L.S., Cowing, J.A. & Davies, W.L. 2009. Evolution and spectral tuning of visual pigments in birds and mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**: 2941–2955.

Hunt, D.M., Dulai, K.S., Cowing, J.A., Julliot, C., Mollon, J.D., Bowmaker, J.K., *et al.* 1998. Molecular evolution of trichromacy in primates. *Vision Res.* **38**: 3299–3306.

Jacob, F. 1977. Evolution and Tinkering. *Science* **196**: 1161-1166.

Joll, L. & Caputi, N. 1995. Geographic variation in the reproductive cycle of the saucer scallop, Amusium balloti, (Bernardi, 1861) (Mollusca:Pectinidae), along the Western Australian coast. *Mar. Freshwater Res.* **46**: 779–792.

Kennedy, D. 1960. Neural Photoreception in a Lamellibranch Mollusc. *The Journal of General Physiology*. 277-299.

Knowles, D.G. & McLysaght, A. 2009. Recent de novo origin of human protein-coding genes. *Genome Res.* **19**: 1752–1759.

Kojima, D. 1997. A Novel Go-mediated Phototransduction Cascade in Scallop Visual Cells. *Journal of Biological Chemistry* **272**: 22979–22982.

Koyanagi, M. & Terakita, A. 2008. Gq-coupled Rhodopsin Subfamily Composed of Invertebrate Visual Pigment and Melanopsin. *Photochemistry and Photobiology* **84**: 1024–1030.

Kozmik, Z., Ruzickova, J., Jonasova, K., Matsumoto, Y., Vopalensky, P., Kozmikova, I., *et al.* 2008. Assembly of the cnidarian camera-type eye from vertebrate-like components. *Proc. Natl. Acad. Sci. U.S.A.* **105**: 8989–8993.

Kraemer, A.C., Serb, J.M. & Adams, D.C. 2015. Batesian mimics influence the evolution of conspicuousness in an aposematic salamander. *J. Evol. Biol.* **28**: 1016–1023.

Land, M. 1966. Activity in the Optic Nerve of Pecten Maximus in Response to Changes in Light Intensity, and to Pattern and Movement in the Optical Environment. *J. Exp. Biol.* **45**: 83-99.

Land, M.F. 1965. Image formation by a concave reflector in the eye of the scallop, Pecten maximus. *J. Physiol. (Lond.)* **179**: 138–153.

Lee, M.G. & Van der Ploeg, L.H. 1987. Frequent independent duplicative transpositions activate a single VSG gene. *Mol. Cell. Biol.* **7**: 357–364.

Lesser, M.P., Carleton, K.L., Bottger, S.A., Barry, T.M. & Walker, C.W. 2011. Sea urchin tube feet are photosensory organs that express a rhabdomeric-like opsin and PAX6. *Proceedings of the Royal Society B: Biological Sciences* **278**: 3371–3379.

Lynch, M. & Force, A. 2000. The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics*. **154**: 459-473.

Malkowsky, Y. & Götze, M.-C. 2014. Impact of habitat and life trait on character evolution of pallial eyes in Pectinidae (Mollusca: bivalvia). *Org Divers Evol* **14**: 173–185.

Malkowsky, Y. & Jochum, A. 2014. Three-dimensional reconstructions of pallial eyes in Pectinidae (Mollusca: Bivalvia). *Acta Zool* **96**: 167–173.

Matheu, A., Pantoja, C., Efeyan, A., Criado, L.M., Martin-Caballero, J., Flores, J.M., *et al.* 2004. Increased gene dosage of Ink4a/Arf results in cancer resistance and normal aging. *Genes & Development* **18**: 2736–2746.

McCormick, M.I. & Manassa, R. 2007. Predation risk assessment by olfactory and visual cues in a coral reef fish. *Coral Reefs* **27**: 105–113.

McCulloch, R., Rudenko, G. & Borst, P. 1997. Gene conversions mediating antigenic variation in Trypanosoma brucei can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Mol. Cell. Biol.* **17**: 833–843.

Meredith, R.W., Gatesy, J., Emerling, C.A., York, V.M. & Springer, M.S. 2013. Rod Monochromacy and the Coevolution of Cetacean Retinal Opsins. *PLoS Genet.* **9**: e1003432.

Minchin, D. 2003. Introductions: some biological and ecological characteristics of scallops. *Aquatic Living Resources*, doi: 10.1016/j.aquliv.2003.07.004.

Mohamed, A. 2003. Structure of the eye and photoreceptors of the nematode Mermis nigrescens.

Mohamed, A.K., Burr, C. & Burr, A.H.J. 2007. Unique Two-Photoreceptor Scanning Eye of the Nematode Mermis nigrescens. *Biol. Bull.* **212**: 206-221.

Ohno, S. 1970. *Evolution by gene duplication.* London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.

Piatigorsky, J. & Wistow, G. 1991. The Recruitment of Crystallins - New Functions Precede Gene Duplication. *Science* **252**: 1078–1079.

Porter, M.L., Blasic, J.R., Bok, M.J., Cameron, E.G., Pringle, T., Cronin, T.W., *et al.* 2012. Shedding new light on opsin evolution. *Proc. R. Soc. B.* **279**: 3-14.

Porter, M.L., Cronin, T.W., McClellan, D.A. & Crandall, K.A. 2006. Molecular Characterization of Crustacean Visual Pigments and the Evolution of Pancrustacean Opsins. *Mol. Biol. Evol.* **24**: 253–268.

Provencio, I., Jiang, G., De Grip, W.J., Hayes, W.P. & Rollag, M.D. 1998. Melanopsin: An opsin in melanophores, brain, and eye. *Proc. R. Soc. B.* **95**: 340-345.

Purschke, G., Arendt, D., Hausen, H. & Muller, M. 2006. Photoreceptor cells and eyes in Annelida. *Arthropod Structure & Development* **35**: 211–230.

Qian, W., Liao, B.-Y., Chang, A.Y.-F. & Zhang, J. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in Genetics* **26**: 425–430.

Ramirez, M.D., Pairett, A.N., Pankey, M.S., Serb, J.M., Speiser, D.I., Swafford, A.J., et al. 2016. The last common ancestor of bilaterian animals possessed at least 7 opsins. doi:10.1101/052902.

Sakmar, T.P. 1998. Rhodopsin: A prototypical G protein-coupled receptor. *Prog. Nucleic Acid Res. Mol. Biol.* **59**: 1–34.

Serb, J.M., Porath-Krause, A.J. & Pairett, A.N. 2013a. Uncovering a Gene Duplication of the Photoreceptive Protein, Opsin, in Scallops (Bivalvia: Pectinidae). *Integrative and Comparative Biology* **53**: 68–77.

Serb, J.M., Porath-Krause, A.J. & Pairett, A.N. 2013b. Uncovering a Gene Duplication of the Photoreceptive Protein, Opsin, in Scallops (Bivalvia: Pectinidae). *Integrative and Comparative Biology*, doi: 10.1093/icb/ict063.

Sivasundar, A. & Palumbi, S.R. 2010. Parallel amino acid replacements in the rhodopsins of the rockfishes (Sebastes spp.) associated with shifts in habitat depth. *J. Evol. Biol.* **23**: 1159–1169.

Smith, A.R. & Carleton, K.L. 2010. Allelic Variation in Malawi Cichlid Opsins: A Tale of Two Genera. *J Mol Evol* **70**: 593–604.

Smith, D.P., Stamnes, M.A. & Zuker, C.S. 1991. Signal Transduction in the Visual System of Drosophila. *Annu. Rev. Cell. Biol.* **7**: 161–190.

Spady, T.C. 2005. Adaptive Molecular Evolution in the Opsin Genes of Rapidly Speciating Cichlid Species. *Mol. Biol. Evol.* **22**: 1412–1422.

Spady, T.C., Parry, J.W.L., Robinson, P.R., Hunt, D.M., Bowmaker, J.K. & Carleton, K.L. 2006. Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays. *Mol. Biol. Evol.* **23**: 1538–1547.

Spaethe, J. 2004. Early Duplication and Functional Diversification of the Opsin Gene Family in Insects. *Mol. Biol. Evol.* **21**: 1583–1594.

Speiser, D.I. & Johnsen, S. 2008. Comparative morphology of the concave mirror eyes of scallops (Pectinoidea). *Amer. Malac. Bull*. **26**: 27–33.

Speiser, D.I., Loew, E.R. & Johnsen, S. 2011. Spectral sensitivity of the concave mirror eyes of scallops: potential influences of habitat, self-screening and longitudinal chromatic aberration. *J. Exp. Biol.* **214**: 422–431.

Stanley, S.M. 1970. *Relation of Shell Form to Life Habits of the Bivalvia (Mollusca)*. Boulder, CO. The Geological Society of America.

Sugawara, T., Terai, Y., Imai, H., Turner, G.F., Koblmüller, S., Sturmbauer, C., *et al.* 2005. Parallelism of amino acid changes at the RH1 affecting spectral sensitivity among deep-water cichlids from Lakes Tanganyika and Malawi. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 5448–5453.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., *et al.* 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* **207**: 454–467.

Terai, Y., Seehausen, O., Sasaki, T., Takahashi, K., Mizoiri, S., Sugawara, T., *et al.* 2006. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol.* **4**: e433.

Vallejo-Marin, M., Buggs, R.J.A., Cooley, A.M. & Puzey, J.R. 2015. Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species Mimulus peregrinus. *Evolution* **69**: 1487–1500.

van Breugel, F., Riffell, J., Fairhall, A. & Dickinson, M.H. 2015. Mosquitoes Use Vision to Associate Odor Plumes with Thermal Targets. *Curr. Biol.* **25**: 1-7.

Wagner, H.-J., Douglas, R.H., Frank, T.M., Roberts, N.W. & Partridge, J.C. 2009. A novel vertebrate eye using both refractive and reflective optics. *Curr. Biol.* **19**: 108–114.

Wang, D., Oakley, T., Mower, J., Shimmin, L.C., Yim, S., Honeycutt, R.L., *et al.* 2004. Molecular evolution of bat color vision genes. *Mol. Biol. Evol.* **21**: 295–302.

Webster, T.P., Selander, R.K. & Yang, S.Y. 1972. Genetic Variability and Similarity in the Anolis Lizards of Bimini. *Evolution* **26**: 523.

Wiltschko, W. & Wiltschko, R. 1991. Magnetic Orientation and Celestial Cues in Migratory Orientation. In: *Orientation in Birds*, pp. 16–37.

Xiao, H., Jiang, N., Schaffner, E. & Stockinger, E.J. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**: 1527-1530.

Yang, S.Y. & Patton, J.L. 1981. Genic Variability and Differentiation in the Galapagos Finches on JSTOR. *The Auk* **98**: 230-242.

Yokoyama, S. 2002. Molecular evolution of color vision in vertebrates. *Gene* **300**: 69–78.

Yokoyama, S. 2000. Molecular evolution of vertebrate visual pigments. *Prog Retin Eye Res* **19**: 385–419.

Yokoyama, S., Tada, T., Zhang, H. & Britt, L. 2008. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **105**: 13480–13485.

Zhang, J. 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol. (Amst.)* **18**: 292–298.

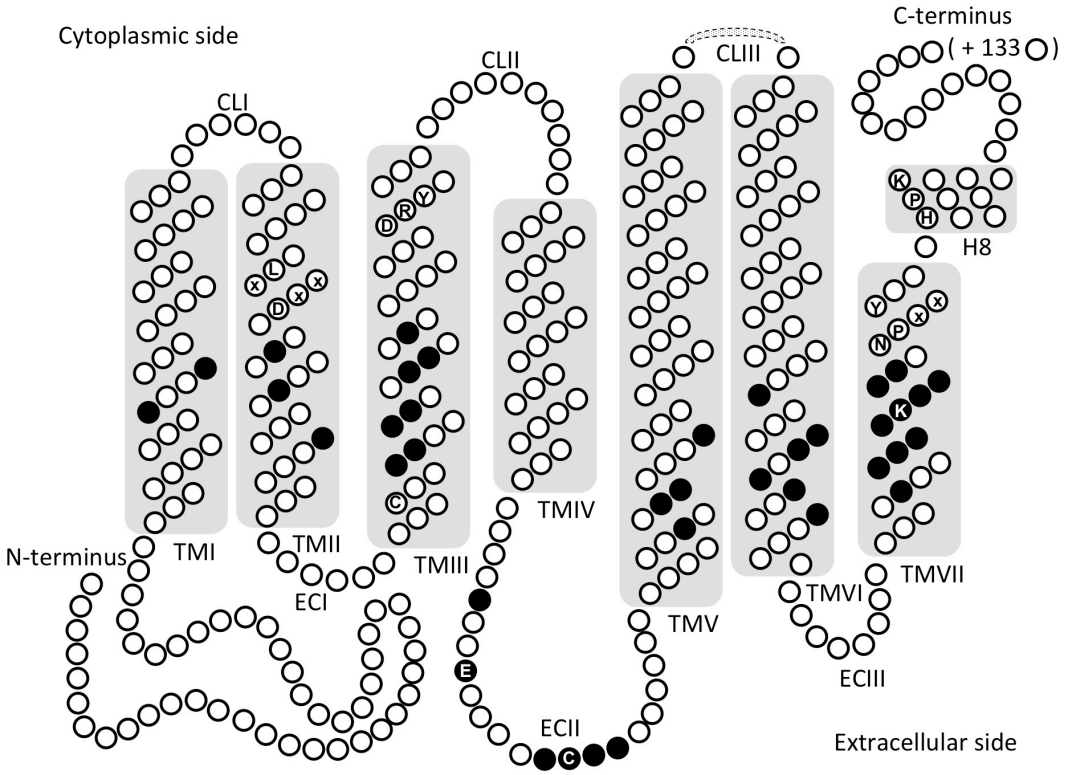
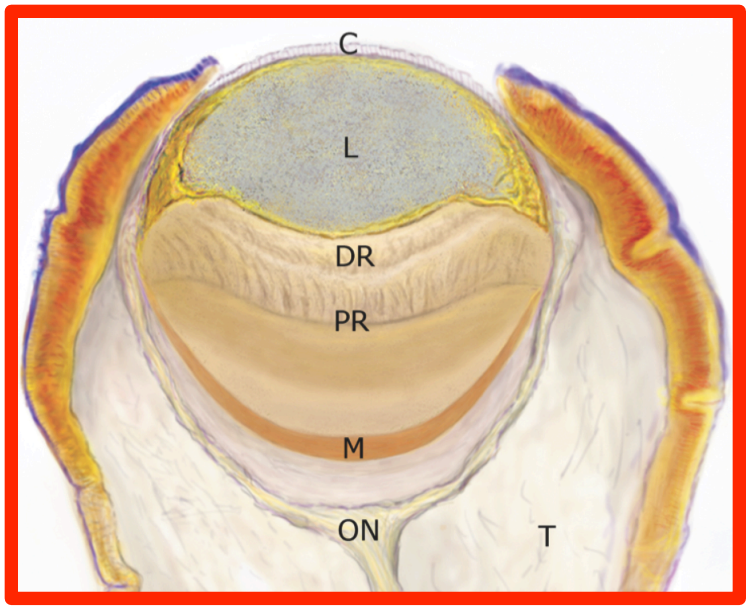


Figure 1-1 Topographical model of G_q-opsin. Each circle represents one amino acid. The greyed rectangles represent alpha helices that transverse the cell membrane. Amino acids predicted to be located inside the G-protein coupled receptor (GPCR) seven transmembrane (TM) regions are numbered TMI through TMVII. H8 represents the predicted eighth helix, which may function as the recognition mechanism for specific G-protein partners. EC and CL indicate the extracellular and cytoplasmic loops, respectively. Solid black circles represent the 38 positions putatively interacting with the chromophore identified by Sekharan *et al*. (2012). LxxxD (TMII), DRY (TMIII), and NPxxY (TMVII) are conserved amino acid motifs important for GPCR functionality. HPK (H8) is a motif conserved in invertebrate bilaterians (Kozmik *et al*., 2008). C (TMIII) and C (ECII) form a disulfide bond, E (ECII) is counterion in some invertebrate opsins (Porter *et al.*, 2012), and K (TMVII) is a chromophore binding site. The C-terminus is truncated by 133 amino acids.



C = Cornea
L = Lens
DR = Distal retina
PR = Proximal retina
M = Mirror
ON = Optical nerve
T = Eye stalk

Figure 1-2 A sagittal cross-section of the scallop eye showing all ocular components present at the adult stage. Pigmented epithelium (red with purple lining) surrounds the right and left perimeter of the eye. *C* cornea, *L* lens, *DR* distal retina, *PR* proximal retina, *M* mirror, *ON* optical nerve, *T* eye stalk.

CHAPTER II

STRUCTURAL DIFFERENCES AND DIFFERENTIAL EXPRESSION AMONG RHABDOMERIC OPSINS REVEAL FUNCTIONAL CHANGE AFTER GENE DUPLICATION IN THE BAY SCALLOP, *ARGOPECTEN IRRADIANS* (PECTINIDAE)

Anita J. Porath-Krause[1, 5], Autum N. Pairett[1, 5], Davide Faggionato[1], Bhagyashree S. Birla[2, 4], Kannan Sankar[3, 4], Jeanne M. Serb[1]

[1]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA USA

[2]Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA USA

[3]Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, IA USA

[4] Interdepartmental Graduate Program in Bioinformatics and Computational Biology, Iowa State University, Ames, IA USA

[5] These authors contributed equally.

Keywords: Rhabdomeric photoreceptor, r-opsin, gene duplication, melanopsin, rhodopsin, vision

Abstract

Background: Opsins are the only class of proteins used for light perception in image-forming eyes. Gene duplication and subsequent functional divergence of opsins have played an important role in expanding photoreceptive capabilities of organisms by altering what wavelengths of light are absorbed by photoreceptors (spectral tuning). However, new opsin copies may also acquire novel function or subdivide ancestral functions through changes to temporal, spatial or the level of gene expression. Here, we test how opsin gene copies diversify in function and evolutionary fate by characterizing four rhabdomeric (G_q-protein coupled) opsins in the scallop, *Argopecten irradians*, identified from tissue-specific transcriptomes.

Results: Under a phylogenetic analysis, we recovered a pattern consistent with two rounds of duplication that generated the genetic diversity of scallop G_q-opsins. We found strong support for differential expression of paralogous G_q-opsins across ocular and extra-ocular photosensitive tissues, suggesting that scallop G_q-opsins are used in different biological contexts due to molecular alternations outside and within the protein-coding regions. Finally, we used available protein models to predict which amino acid residues interact with the light-absorbing chromophore. Variation in these residues suggests that the four G_q-opsin paralogs absorb different wavelengths of light.

Conclusions: Our results uncover novel genetic and functional diversity in the light-sensing structures of the scallop, demonstrating the complicated nature of G_q-opsin diversification after gene duplication. Our results highlight a change in the nearly ubiquitous shadow response in molluscs to a narrowed functional specificity for visual processes in the eyed scallop. Our findings provide a starting point to study how gene duplication may coincide with eye evolution, and more specifically, different ways neofunctionalization of G_q-opsins may occur.

Background

Organisms detect environmental stimuli using an array of sensory receptors. Changes to the genetic basis of these sensory receptors has been shown to allow organisms to exploit new ecological niches [1] or alter signaling between conspecifics [2], which can affect individual fitness and, ultimately, have evolutionary consequences for the species. Duplication of the genes that code for the sensory receptor proteins is thought to play an important role in expanding the diversity of sensory systems by providing new genetic material for novel phenotypes [3–6]. If gene duplicates are retained, they can follow one of three evolutionary fates (first outlined by [7]; see also expanded models reviewed by [8–10]). First, if both paralogs have the exact same function or suite of functions, the existence of a second copy can increase production levels of encoded protein ("gene conservation" [11]). Under this scenario, the second copy provides functional redundancy that can buffer against neutral loss-of-function mutations over evolutionary time. However, more dramatic functional divergence may occur following the duplication event. In the second scenario, if the original gene managed a suite of functions, such as enzymatic activity and signal transduction, the duplicated copies could subdivide these tasks ("subfunctionalization" [12]). Subfunctionalization of paralogs may include changes in spatial or temporal expression patterns [13] and may release one gene copy from adaptive constraint ("escape from adaptive conflict" model [14]) so that both copies can be optimized for particular tasks [15]. Finally, one copy of the duplicated gene can acquire a novel function while the other copy retains the original, pre-duplication function ("neofunctionalization" [7]).

In photosensory systems, the ability of an animal to become sensitive to a broader range of wavelengths is most often mediated by an increase in the number of opsins [16–22]. Opsins

encode a class of G-protein coupled receptors (GPCRs), proteins with seven alpha-helical domains that transverse the cell membrane (helix, H1-7) interspaced by loops that extend into the cytoplasm (cytoplasmic loops, CL1-3) and outside of the photoreceptive cell (extracellular loops, EC1-3). Opsins covalently bind a light-absorbing vitamin-A derived chromophore, such as 11-*cis*-retinal, using a lysine residue in H7. Together, the opsin protein and chromophore molecule form a photopigment sensitive to a specific portion of the light spectrum. Photopigments are often characterized by the wavelength at which the absorbance of light is the greatest (λ_{max}). When 11-cis retinal absorbs a light photon, it isomerizes to an all-trans state. As a result, the opsin undergoes a conformational change and releases a complex of heterotrimeric guanine nucleotide-binding proteins (G-proteins), which are specific to that opsin (reviewed in [23]). The dissociated alpha-subunit of the G-protein activates the phototransduction cascade through second messenger molecules. Depending on the particular transduction pathway initiated by opsin, the photoreceptor cell may either hyperpolarize (e.g., G_t-protein coupled opsins in ciliary cells) or depolarize (e.g., G_q-protein coupled opsins in rhabdomeric cells) [24]. Opsin specificity to its G-protein partner is regulated by G-protein binding sites [25] and is associated with particular amino acid motifs in the fourth cytoplasmic loop [26]. Phylogenetically, opsins group into clades based, in part, by the G-protein partner and to a lesser extent by photoreceptor type (rhabdomeric versus ciliary cells) [27, 28].

Because a photopigment can only absorb a portion of the light spectrum, increasing the number and diversity of opsins through gene duplication and divergence allows an expansion of the photoresponse to new wavelengths of light. This may lead to color discrimination, if the photopigments have different light sensitivities. Under this neofunctionalization model, changes

in the amino acid residues at positions that interact with the chromophore (e.g., "spectral tuning sites") shift the wavelength at which absorbance is the greatest (λ_{max}) of the duplicated visual pigment. Thus, the potential advantages for organisms with multiple and genetically diverse photopigments include extending the range of spectral perception, new functionality under different light conditions, generation of wavelength-specific behaviors, or providing the molecular substrate in the retina for color vision (reviewed in [29]). Any of these phenotypes may allow an animal to occupy new or more heterogeneous photic niches [30, 31].

While it is well-documented that duplicated opsin genes most often attain a new λ_{max} by neofunctionalization [32–40] it is less understood what other phenotypic outcomes may follow the duplication of opsin genes (but see [21]). Photoreceptors in invertebrates occur in multiple tissue types and in different life stages, and can function as both ocular and extra-ocular sensory receptors [41–46]. Thus, in invertebrates, neofunctionalization of opsins may include co-option between tissues, organs, or life stages after a gene duplication event. In order to distinguish among different evolutionary outcomes of opsin duplication and what effect gene duplication may have in the evolution of the photoreceptive cells and organs in a given system [47], it is necessary to first identify and then characterize the diversity of opsin proteins that are present.

Here, we assess the evolutionary history of G_q-opsins in scallop to examine the role of gene duplication in producing extant diversity. The molecular basis of photoreception in the scallop is complex. The mirror-type eyes of scallops contain at least two different phototransduction systems based on opsins that presumably couple with G_o- and G_q-proteins [48]. Previously, we identified a duplication event of scallop G_q-protein coupled opsins that occurred over 230 Mya

[49]. Because gene copies with identical gene function are unlikely to be maintained in the genome unless the new duplicate is advantageous [50], the long-term retention of these opsin duplicates in the scallop lineage suggests a fitness cost if the copies are not maintained. For these duplicates to persist over evolutionary time, opsin copies must have diverged phenotypically under one or more of the evolutionary fate models described above. To test this hypothesis, we determined the evolutionary fates of these duplicated scallop opsins. We first captured the genetic diversity of G_q-protein coupled opsin genes (herein *opnGq* for the gene or the coding region, and OPNGq for the protein) by generating transcriptomes of photosensitive tissues from adult animals and placed the genetic diversity of scallop G_q-opsins into an evolutionary framework by employing a phylogenetic analysis. We next asked how might these scallop OPNGq proteins interact with a chromophore. To do so, we capitalized on the x-ray crystallography data from the squid OPNGq ("squid rhodopsin") [51, 52] to model the tertiary structure of the scallop OPNGqs. Then, we examined if the protein characteristics of each paralog differ. As a first approximation to identify differences in λ_{max} among scallop G_q-opsins, we leveraged existing computational models that estimate electrostatic interactions between the amino acids and the chromophore of squid OPNGq and applied them to the scallop data. Finally, we examined differences in gene expression of *opnGq* paralogs across both ocular and extra-ocular photoreceptive organs. From these lines of evidence, we show that scallop G_q-opsin paralogs differ in 1) the biochemical properties of amino acid residues interacting with the chromophore; 2) expression levels of the gene; and 3) spatial expression of the gene among light-sensitive tissues in the adult organisms.

Methods

Transcriptome assembly and gene analyses

Thirty-six adult individuals of the bay scallop, *Argopecten irradians* (Pectinidae), were collected from the Gulf of Mexico near Sanibel, Florida during July, 2012. The adults were kept in recirculating saltwater tanks under a light regime of 13 hours of light and 11 hours of dark per 24-hour cycle. To maximize the likelihood of capturing all G_q-opsin transcripts expressed, we collected tissues under both light and dark treatments (nine hours of light vs. nine hours of dark), with the expectation that the highest level of opsin expression would occur nine hours after sunrise [53, 54]. The tissues from dark-treated scallops were dissected under red-light. All eyes from the left and right mantles were collected and pooled for each animal (~60 eyes/individual). Small sections of mantle tissue were sampled along the anterior-posterior axis from both left and right valves and pooled for each individual. A portion of adductor muscle equivalent in volume to the dissected eye tissue was collected from each individual. RNA was extracted from the three tissue types using the Ambion RiboPure RNA extraction kit (Life Technologies). RNA samples from the tissues of one light-treated and one dark-treated individual were sent to the Iowa State University DNA Facility for library creation and transcriptome sequencing on an Illumina HiSeq2000. Nearly 1.5 trillion 100 base pair (bp) paired-end reads were generated from six libraries: light/dark eyes, light/dark mantle, and light/dark adductor. A *de novo* assembly of a reference transcriptome from all six libraries was created in the Trinity sequence assembly and analysis pipeline [55] by first normalizing the raw reads to remove redundancy with the Trimmomatic script, then assembling the quality trimmed reads. This assembly resulted in 231,391 transcripts with a contig N50 of 2078 and an average contig length of 971 bp. The assembled transcriptome data was given the reference name of "AirradFL." Opsin sequences

from two other scallop species [56] were used as queries to identify G_q-opsin sequences in the AirradFL reference transcriptome using BLAST. Putative opsin sequences from the AirradFL reference transcriptomes were blasted back to the NCBI nonredundant (nr) database to further confirm the sequence identities. Gene and protein nomenclature follows the general guidelines in invertebrate model organisms (e.g., www.wormbase.org), where gene and transcript names (italicized) are composed of a three-letter species prefix, followed by a hyphen, the class (homolog) of the gene, and a number (e.g., *Air-opnGq1*). The number provides the order of gene discovery of paralogs within a species or lineage. Proteins use the gene name, with the gene abbreviation without italics and in all uppercase (e.g., Air-OPNGq1).

Phylogenetic analysis

To determine the phylogenetic placement of putative scallop G_q-opsins, we compiled G_q-opsin sequences from genomes, transcriptomes or single genes from public databases at Genbank (www.ncbi.nlm.nih.gov/genbank/) and assembled data from Porter *et al*. [27] (Additional file 2: Table 2-S2). We queried all five publically-available molluscan genomes for additional G_q-opsins: pearl oyster, *Pinctada fucata* (June, 2013); Pacific oyster, *Crassostrea gigas* (June, 2013); freshwater snail, *Biomphalara glabrata* (June, 2013); owl limpet, *Lottia gigantea* (June, 2013); and sea hare, *Aplysia californica* (June, 2013). G_q-opsin sequences were found by blasting scallop opsins against predicted gene models from each molluscan genome using tblastx and an E-value cutoff of 1e-3. When gene models were not available, the genome contigs/scaffolds were used. The putative G_q-opsins identified through BLAST were then reciprocally blasted back to the NCBI nonredundant (nr) database and subjected to phylogenetic analyses with known metazoan G_q-opsins to confirm their identity.

Amino acid sequences of the 96 opsins from 42 taxa, including four annelids, 38 arthropods, 21 molluscs, and six platyhelminthes, (Additional file 2: Table 2-S2) were aligned using MAFFT v 7.017 [57] as implemented in Geneious (v5.6.7). (www.biomatters.com/geneious). This dataset included opsins from the G_i- and G_o-opsin families to test the monophyly of the G_q-opsin clade. The G_o-opsin from *Argopecten irradians* was used to root the phylogeny. The aligned dataset was then manually trimmed to remove long C- and N- terminus sequences and remove a single large (>50 aa) gap around position 258 in the H6. The trimmed, aligned dataset contained 355 amino acids. The best-fit model of protein evolution for this dataset was determined using ProtTest [58], which found the LG+G+I+F model [59] to have the lowest Akaike Information Criteria score (AIC). A maximum likelihood (ML) phylogeny of the aligned dataset was constructed using Randomized Axelerated Maximum Likelihood (RAxML) v 8 [60]. Node support was calculated using 1000 rapid bootstrap replications as implemented in RAxML. Using the same model of protein evolution, we also analyzed the data under Bayesian inference using MrBayes v3.2.6 [61] on the XSEDE tool available through the CIPRES Science Gateway [62]. We used the Metropolis Coupled Markov Chain Monte Carlo method with one cold and three hot chains for 3.1 million generations with a burnin of 1000 for two independent runs. Convergence was determined when the potential scale reduction factor (PSRF) approached 1.

PCR confirmation of scallop opsin transcripts

All opsin transcripts were confirmed to be single genes by PCR amplification of the complete coding region with UTR-specific primers from both cDNA and genomic DNA (Qiagen DNeasy Blood and Tissue kit) (Additional file 1: Table 2-S1). PCR products were size-screened using

agarose gel electrophoresis, bands of expected size were gel extracted (Qiagen Qiaquick Gel Extraction kit) and cloned using chemically competent *E. coli* cells (TOPO TA Cloning Kit with pCR2.1-TOPO). Positive colonies from blue-white screening were Sanger sequenced using an ABI 3730 Capillary Electrophoresis Genetic Analyzer at the Iowa State University DNA Sequencing Facility. The resulting sequences were translated and compared against contigs from the transcriptome. Using the same approach, we confirmed that a large contig sequence containing two G_q-opsin transcripts (*Air-opnGq3* and *Air-opnGq4*) and an intergenic region of ~1690 bp was present in the genome. Because repetitive motifs can indicate gene duplication due to transposable elements [63], we searched for repetitive motifs in this intergenic region. To do so, the nucleotide sequence of the whole contig was screened with the RepeatMasker Web server v open-4.0.5 (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) using the cross_match search engine on slow speed/sensitivity and the bivalves *Crassostrea gigas*, *Pinctada fucata*, and *Mizuhopecten yessoensis* as DNA sources.

Homology modeling of scallop G_q-opsins

To identify amino acid changes that may result in functional differences among scallop G_q-opsins, we compare the Air-OPNGqs to the only molluscan opsin with a resolved crystal structure, the *Todarodes pacificus* "rhodopsin" (Tpa-OPSGq1; Genbank accession X70498) [51] We followed the amino acid numbering system of the squid where the first amino acid position in our alignment begins with the start codon (Met) of Tpa-OPNGq1. To examine the degree of resemblance among protein sequences, we calculated pairwise percent similarity of the scallop and squid amino acid sequences in the BLASTP 3.2.1 [64, 65] at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins).

We also used the protein alignment to identify amino acid residues that may interact with the chromophore. We applied a quantum mechanics/molecular mechanics model based on the crystal structure of Tpa-OPNGq1 [66], which predicts the involvement of 38 sites in spectral tuning of G_q-opsins. We examined differences in the Air-OPNGq and Tpa-OPNGq1 sequences at these sites and noted changes in the biochemical properties of the residues.

Next we employed bioinformatic homology modeling to predict the tertiary structure of the four scallop G_q-opsin proteins. These models were based on the template of the only available crystal structure for a G_q-opsin, the rhodopsin from squid *Todarodes pacificus* 2ZIY [52]. The tertiary structure models of four scallop opsins (Air-OPNGq1, Air-OPNGq2, Air-OPNGq3 and Air-OPNGq4) were predicted using the Iterative Threading ASSEmbly Refinement (I-TASSER) server [67, 68]. The squid 2ZIY template was used to retrieve model proteins of similar folds from the Protein Data Bank (PDB) library using a locally installed meta-threading library. The continuous fragments excised from PDB templates were re-assembled into full-length models by replica- exchange Monte Carlo simulations and the unaligned regions were built by *ab-initio* modeling. The structure was then further refined with a second fragment assembly simulation. No restraints such as inter-residue contacts or inter-residue distances were specified for the modeling. For each G_q-opsin, the top five predicted structures from I-TASSER were used for further quality assessment.

Assessing the quality of the modeled tertiary structures.

The quality of the modeled structures was assessed using the Ramachandran plot and the confidence score (C-score) (Additional file 5: Table 2-S3) from the I-TASSER server. The Ramachandran plot is a graph of the backbone dihedral angles ψ against ϕ of the amino acid residues in the structure. Good quality models have more than 90% of the residues in allowed regions (i.e. most favored and additionally allowed regions) of the Ramachandran plot. The Ramachandran plot of the modeled structures was obtained using PROCHECK [69] which has been implemented as part of the PDBSum Server [70].

The C-score (from I-TASSER server) is a scoring function to rank models based on their quality and is defined using the significance of threading template alignments and the convergence parameters of the structure assembly simulations (for more details see [67]). C-scores are typically between -5 and 2 with higher values representing better models. However it has been observed that the C-score is particularly low (and negative) for membrane proteins. The "best" models of the four G_q-opsin sequences were selected based on the highest C-score and maximum percentage of residues in the most favored and generously allowed regions according to the Ramachandran plots.

To quantify the overall shape differences among G_q-opsin tertiary structures, we performed a whole-molecule comparison between the predicted tertiary models calculating the Root-Mean-Square Deviation (RMSD) of the atomic positions of the alpha carbons between one opsin against each other. RMSD provided a quantitative computation of the average distance between the backbone atoms of two superimposed proteins. Variation in Air-OPNGq sequence length did

not impact the RMSD values because a small portion of the N- and most of the C-termini were truncated from each sequence so the comparison occurs only between superimposed atoms. For RMSD comparison, only common one-to-one aligned residues, were included (V19 to K342). The values between each pair of structures were calculated using the standard 'align' program in PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). Lower RMSD values indicate a higher similarity between structures.

Scallop gene expression data

Paired-end RNA-seq data for three scallop tissues (eye, mantle, and adductor muscle) from the light treatment were aligned against the AirradFL assembly (nonredundant set of 231,391 transcripts grouped into 176,417 "genes") using Bowtie v. 1.0.1 [71] followed by read abundance estimation with RSEM v. 1.2.9 [72] through the Trinity sequence and assembly pipeline v. 2013_2-25 [55]. Relative levels of expression in Fragments Per Kilobase per Million fragments mapped (FPKM) for a given transcript were calculated using the Trinity toolkit v. 2013_2-25 [55]. We accepted expression levels for a given transcript when the FPKM value was equal to or greater than one as a conservative approach to compare levels of relative expression among tissue types. Because tissues under the light treatment had the greatest levels of G_q-opsin expression, only the results from light-treated tissues are reported here.

Oyster gene expression data

To compare interspecific differences in G_q-opsin expression patterns between bivalve taxa, opsin gene expression data for the Pacific oyster, *Crassostrea gigas*, were mined from the oyster genome database (OysterBase, www.oysterdb.com). We identified *opnGqs* from oyster by

blasting our scallop G_q-opsins against the database using the OysterBase blast tools with default settings. Gene expression data in RPKM (Reads Per Kilobase per Million) of the oyster G_q-opsins (*Cgi-opnGqs*) were curated for each adult tissue type (digestive gland, gills, gonad, hemolymph, labial palp, mantle, and pallial mantle) and larval life stages (trochophore, D-shape larva, umbo larva, and pediveliger) from the website (OysterBase, www.oysterdb.com) and supplementary data tables (Table S12, S14) in Zhang [73]. However, comparing gene expression changes between the oyster (in RPKM) and scallop (in FPKM) tissues could only be described in relative terms.

Results

Transcriptomic and phylogenetic analyses reveal four G_q-opsin genes in scallop

To determine the number of G_q-opsin genes in scallop, we performed deep transcriptome sequencing of tissue-specific libraries derived from dissected eyes, mantle tissue, and adductor muscle of *Argopecten irradians*. From light and dark treated animals, four transcripts were identified as putative *opnGqs* using a similarity-based analysis pipeline described in Pairett and Serb [56], which we named *Air-opnGq1*, *Air-opnGq2*, *Air-opnGq3*, and *Air-opnGq4* with ascending numbering according to the history of discovery (GenBank accession numbers KT426908, KT426909 KT426910, and KT426911). Visual inspection of the back mapped reads to each identified G_q-opsin sequence did not show any obvious misassembled regions or mismatches. The proteins varied in amino acid percent similarity (the ratio of residues with similar physio-chemical properties shared between two sequences), which were the greatest between Air-OPNGq2 and Air-OPNSGq3 at 80.9%, and lowest between Air-OPNGq1 and Air-OPNGq4 (72.9%) (Table 2-1). Amino acid percent similarity was more conserved between the

aligned Helix 1 (H1) through H7, and ranged from 92.6% (Air-OPNGq2 versus Air-OPNGq3) to 76.9% (Air-OPNGq1 versus Air-OPNGq4) (Table 2-1). Transcripts also differed in the sequence length from the first Met codon to the beginning of H1 (35-49 amino acids) and between the end of H7 and the stop codon (135-184 amino acids) (Fig. 2-1; Table 2-2).

To determine how Air-OPNGqs were evolutionarily related to other G_q-opsins, we conducted a phylogenetic analysis of their translated amino acid sequences with 96 metazoan opsins (Additional file 2: Table 2-S2). Under both maximum likelihood and Bayesian inference, all four scallop sequences belonged to a clade that included G_q-opsins from four other bivalve species: two oysters (*Pinctada fucata*, *Crassostrea gigas*) and two additional scallops (*Placopecten magellanicus*, *Mizuhopecten yessoensis*) (Fig. 2-2, green box). Within this clade, there was one difference between the ML and BI topologies, where ML placed the two oyster OPNGq1s as the sister group to the scallop G_q-opsins 2-4, and the BI topology placed all bivalve OPNGq1s in a single clade (grey box in Additional file 3: Fig. 2-S1). However, values supporting these relationships were low (47% bootstrap support; 54 posterior probability). The bivalve-specific G_q-opsin clade (OPNGq1-4) was the sister group to a clade of opsins from cephalopod and gastropod molluscs, and part of a larger clade of well-characterized vertebrate (e.g., melanopsin) and arthropod (e.g., *Drosophila* rhodopsin) G_q-opsins (Fig. 2-2). A second molluscan G_q-opsin clade was also recovered which contained oyster and gastropod opsins, but no scallop opsins (Fig. 2-2, red box). A complete, uncollapsed ML phylogram is available as a supplemental document (Additional file 4: Fig. 2-S2)

We then asked whether the four scallop G_q-opsins possess the specific amino acid residues and sequence motifs required for photosensitivity. In addition to the seven transmembrane α-helices, it has been experimentally demonstrated that G_q-opsin proteins require certain sequence motifs to maintain structural integrity and bind to the chromophore [74]. These include: 1) two Cys residues in the TM3 and EC2 domains that are involved in disulfide bond formation, 2) a Glu180 in the EC2 that functions as a counter ion to the positive charge of the protonated Schiff base [75], 3) a E/DRY motif near the TM3/CL2 boundary that helps stabilize the inactive-state conformation [76], 4) Asn87 and Tyr111 residues that are hydrogen binding partners for the protonated Schiff base [52], 5) a lysine residue in TM7 that is covalently linked to the chromophore, and 6) a conserved NPxxY motif in the TM7 [74]. We found that all four scallop proteins were invariant for the expected amino acid residues and motifs needed for correct conformation with the exception of the E/DRY motif (Table 2-2). This motif was variable among the scallop opsins, where Y134C in Air-OPNGq2 and Y134F in Air-OPNGq3 and Air-OPNGq4. In addition, we examined a motif (positions 319-321) in the fourth cytoplasmic loop, which has been experimentally demonstrated to be important for opsin-G_t-protein interactions (positions 310-312 in bovine rhodopsin) [25]. Three of the four scallop opsins contain a HPK motif, an evolutionary conserved sequence that appears to be specific to G_q-protein binding [77] (Table 2-2). Air-OPNGq4 had a HPR motif, but R has similar biochemical properties to K. Based on these data, we conclude that the four transcripts are indeed OPNGqs possessing the amino acid residues required for molecular stabilization, chromophore binding, and G-protein interaction and thus likely form photopigments.

G_q-opsin transcripts are not the result of alternative splicing

To determine whether the four different *opnGq* transcripts were the result of alternative splicing of the same gene, we developed target-specific primers (Additional file 1: Table 2-S1) from the flanking UTR sequences for each *Air-opnGq*. We then compared these sequences derived from genomic DNA (gDNA) to transcripts derived from the transcriptomes. Alignments of 5'- and 3'-UTR DNA sequences and coding regions were identical between the transcripts and gDNA templates (data not shown). The flanking UTR sequences were not conserved and could not be unambiguously aligned across the four *Air-opnGqs* (Additional file 6: Fig. 2-S3).

While three of the four *Air-opnGq* sequences lacked introns, we identified a 393 bp intron within the region coding of H3 that was unique to *Air-opnGq1*. Additionally, gDNA sequencing determined that *Air-opnGq3* and *Air-opnGq4* were located in tandem, but in reverse orientation, with a 1690 bp intergenic region between the two coding regions. No repeat regions or putative transposable elements were identified in the intergenic region (data not shown). Variation in intron pattern and UTR sequences among the G_q-opsins indicates that these four genes are most likely located on different physical places in the genome and are four separate loci.

Predicted tertiary structure and chromophore-associated residues differ among scallop G_q-opsins

We generated three-dimensional models for each Air-OPNGq using crystallography data from the squid "rhodopsin" [52] as a template for homology models. This allowed us to examine differences in the tertiary structure among the four G_q-opsin sequences. The best model for each Air-OPNGq was selected based on the highest C-score and maximum percentage of residues in

the most favored and generously allowed regions according to the Ramachandran plots (Additional file 5: Table 2-S3). To quantify the overall shape differences among G_q-opsin tertiary structures, we performed a whole-molecule comparison between the predicted tertiary models calculating the Root-Mean-Square Deviation (RMSD) of the atomic positions of the alpha carbons between one opsin against each other. Based on the RMSD of atomic values, tertiary structures differed from 0.354 to 0.699 angstroms, where lower RMSD values indicate higher similarity between structures (Table 2-1). Predicted tertiary structures were the most similar among Air-OPNGq1, Air-OPNGq2, and Air-OPNGq3 proteins (RMSD ranged between 0.354 and 0.408), while Air-OPNGq3 was most different from Air-OPNGq4 (RMSD = 0.699) (Table 2-1). Air-OPNGq3 and Air-OPNGq4 are more different in tertiary structure from each other than either are to squid rhodopsin (RMSD = 0.503 and 0.601).

We then examined if the positions predicted to interact with the chromophore differ in their residues among the four scallop G_q-opsins. We employed results from a quantum mechanics/ molecular mechanics (QM/MM) model based on the Tpa-OPNGq1 crystal structure [66]. This model predicts 38 amino acid sites that may play a role in spectral tuning of G_q-opsins. The scallop G_q-opsins differed from the Tpa-OPNGq1 at seven of the 38 positions, but only three of these had residues with another biochemical property (Fig. 2-3, blue dots). Among the four scallop G_q-opsins, seven of the 38 positions varied (Fig. 2-3, red dots). At four positions, at least one of the scallop opsins had an amino acid residue with a different biochemical property. Position 92 was the most divergent among Air-OPNGq proteins and included nonpolar aliphatic/hydrophobic (Air-OPNGq1 and Air-OPNGq2) and aromatic residues (Air-OPNGq3), while Air-OPNGq4 had a positive polar residue (Lys) at this position. At position 275, a

conserved serine was substituted by cysteine in Air-OPNGq4, and at position 306, adjacent to the lysine forming the Schiff base, Air-OPNGq1 and Air-OPNGq4 have an hydrophilic residue instead of an hydrophobic/aliphatic residue (Fig. 2-3).

G_q-opsins are differentially expressed across the eye, mantle and adductor muscle tissues

To determine whether the expression patterns from the four G_q-opsins in *A. irradians* differ spatially, we compared the relative expression level of each G_q-opsin among the six tissue-specific transcriptomes from adult animals collected after a nine-hour light treatment or a nine-hour dark treatment. We found that spatial expression of the four G_q-opsins was consistent in the light and dark adapted animals (data not shown); however, tissues under the light treatment had the greatest levels of G_q-opsin expression and we only the report these results here.

We found all four scallop G_q-opsins were expressed in the eye. Outside of the eye, both *Air-opnGq1* and *Air-opnGq2* were expressed in the mantle, but only *Air-opnGq2* was expressed in the adductor muscle at levels above our expression threshold (≥ 1.0 FPKM; Fig. 2-4). As a general pattern across all tissue types, *Air-opnGq2* had the highest expression levels, while *Air-opnGq4* was expressed at the lowest level or not at all. When comparing relative expression levels in the eye, *Air-opnGq2* and *Air-opnGq3* had the highest relative expression levels with *Air-opnGq2* expression (10,001.27 FPKM) at ~38 times higher than *Air-opnGq3* (260.64 FPKM), 275-times higher compared to *Air-opnGq1* (36.46 FPKM), and over 5800-times higher *Air-opnGq4* (1.72 FPKM) (Fig. 2-4).

We then examined relative levels of gene expression in the Pacific oyster (*Crassostrea gigas*). Since this species is eyeless as an adult, we anticipated that its genome would contain a limited number of G_q-opsins. However, our analyses identified three different G_q-opsins in the *C.*

gigas genome (*Cgi-opnGq1, Cgi-opnGq2A,* and *Cgi-opnGq2B)* that showed a degree of differential expression across tissues and life stages. *Cgi-opnGq1,* the oyster G_q-opsin most closely related to the scallop opsins identified here (Fig. 2-2, green box), was found to have low (<1.0 RPKM) expression levels across the adult oyster tissues, but relatively higher expression in the larval umbo (2.508 RPKM) and pediveliger (21.355 RPKM) stages. *Cgi-opnGq2A* and *Cgi-opnGq2B* belonged to a second clade of gastropod and bivalve G_q-opsins (Fig. 2-2, red box). *Cgi-opnGq2A* was most highly expressed in the adult tissues, with the labial palp (organs that move food to the mouth for ingestion) and pallial mantle (the tissue most similar to the scallop eye-containing mantle edge) showing the greatest *Cgi-opnGq2A* expression (2.290 RPKM and 4.080 RPKM, respectively). *Cgi-opnGq2B* showed the lowest expression across all tissues and life stages (<1.0 RPKM).

Discussion

The duplication of opsin genes is considered to be an important mechanism for the expansion of light-sensing capabilities of photosensory systems by either enhancing wavelength discrimination or increasing the spatial expression. While some of the best studied examples of photosensitivity expansion are the separate origins of color vision in insects [22, 42, 78] and vertebrates [17, 79, 80], where shifts in absorbance spectra are attributed to nonsynonymous substitutions to the coding region of one opsin copy, post-duplication fates of opsins need not be limited to changes in the coding region. Functional divergence of opsin copies can also be driven by changes to the untranslated regions of the gene, which contain regulatory elements influencing gene expression and translation. This latter phenomenon has been less studied in post-duplicated opsins (but see [81]). While we did not directly investigate regulation of scallop

G_q-opsin, our discovery of tissue-specific expression of G_q-opsin paralogs in the scallop, *Argopecten irradians*, not only provides circumstantial evidence that there may be differences in regulatory regions, but offers an opportunity to investigate how these gene copies diversified in function and evolutionary fates. One-to-one matches between transcript and genomic amplicons strongly support the presence of at least four G_q-opsin paralogs in the *A. irradians* genome. All four genes were identified as G_q-opsins by both sequence similarity and phylogenetic analysis, and are most likely the result of duplication events in a lineage that includes the orders Pectinoida and Limoida [49], either through whole genome duplication events [82] or duplication of small segments of the genome [83]. The specific timing of these events will require denser taxonomic sampling within the subclass Pteriomorphia, but if the phylogenetic pattern from our study holds, it would appear that *opnGq1* and *opnGq4* are derived from the first round of gene or genome duplication. Subsequently, *opnGq4* may have undergone a tandem duplication, and the paralog underwent a second round of duplication to create *opnGq2* and *opnGq3* (Fig. 2-5).

We present evidence that all four *Air-opnGqs* products, when reconstituted with the proper chromophore, could form photopigments. Each scallop G_q-opsin has the sequence motifs necessary for protein conformation and chromophore binding (Table 2-2). Tertiary structural models developed for each Air-OPNGq contain the expected protein domains and loops for a functional opsin protein. Interestingly, all four scallop protein models predict eighth and ninth cytoplasmic α-helices (Fig. 2-1), features unique to G_q-opsins [51]. In the Tpa-OPNGq1 crystal structure, the C-terminus of H9 interacts with the cytoplasmic extension of H6, that together with H5 form a rigid column projecting 25Å from the membrane surface; however the rotational

freedom of H9 is restricted by its interactions with H8. Thus, others have predicted that this four-domain cytoplasmic feature, in conjunction with the HKP motif in H8 [26], functions as the recognition mechanism for specific G-protein partners [51]. In summary, our bioinformatic analyses support that all four scallop G_q-opsins form photopigments that could be used to detect light. How might these gene copies have diverged after the duplication event? Molecular changes in paralogous scallop opsin genes appeared to have occurred both outside and within the protein-coding region.

We find differential gene expression across ocular and extra-ocular structures in the adult, suggesting there have been changes in the regulatory regions of scallop G_q-opsin paralogs. Specifically, while all *Air-opnGqs* are expressed in eyes, the level of expression is vastly different (ranging from a 38- to 5815-fold difference). In addition, only two of the four G_q-opsins, *Air-opnGq1* and *Air-opnGq2*, are significantly expressed outside of the eye, and presumably they are used in a nonvisual context such as the "shadow response" [84]. Taken together, these data suggest that scallop opsin paralogs are used in different biological contexts. Some may preferentially be employed in eyes (*Air-opnGq3* and *Air-opnGq4*), while others (*Air-opnGq1* and *Air-opnGq2*) are used for both ocular and extra-ocular based functions.

Spatial patterning and expression level differences among the scallop G_q-opsin paralogs suggest they have undergone neofunctionalization since duplication. When we compare the scallop opsin expression data to the closest related bivalve with a sequenced genome, the Pacific oyster, *Crassostrea gigas* [73] we find a dramatic difference in the relative levels of gene expression and spatial patterning. From the oyster genome, we identified three G_q-opsins, but only one (*Cgi-*

opnGq1) was phylogenetically similar to the scallop opsins (Fig. 2-2). This *Cgi-opnGq1* is broadly expressed at low levels across the adult non-ocular tissues (e.g., 0.10 RPKM in mantle tissue to 0.29 RPKM in gonad) [73]. In contrast, the adult scallop has high levels of expression (up to 10,001.27 FPKM) of different G_q-opsin gene copies in eyes, and low or no expression of these opsins in non-ocular tissues (Fig. 2-4). Could an increase in opsin expression level and/or greater number of gene copies be related to the origin of eyes? Currently available opsin sequences from bivalve species represent a very restricted taxonomic sampling. But based on the nearly ubiquitous shadow response in Bivalvia and Gastropoda, the few instances of eyes in bivalves [85], and the results from our study, we anticipate that the ancestral state for G_q-opsin spatial expression in bivalves is across multiple tissue types while the derived condition of spatial expression is narrowed (limited) to eyes and may indicate functional specificity for visual processes. If one or both of the scallop opsin duplication events were concurrent with the origin of eyes, it would support the notion of neofunctionalization of the new G_q-opsin copies.

Do the differential levels of gene expression indicate an even finer spatial partitioning of *Air-opnGqs*? We anticipate this to be the case. Depending on the scallop species, an adult animal can have between 35 to over 200 eyes along the mantle margins lining both valves (Serb, unpublished) that can vary in size [86, 87]. Visual fields from adjacent eyes overlap such that, as a conservative estimate, at least five eyes would convey similar information from a given point in the environment (estimated from a 30-eyed animal [88]). One way to reduce functional redundancy would be to distribute Air-OPNGq proteins of dissimilar absorbance spectra across non-adjacent eyes. However, due to the limitations of library construction, which required the pooling of all 60 eyes from one light- and 60 eyes from one dark- treated animal, we are unable

to determine if a single eye expresses all or a just subset of *Air-opnGqs*. Furthermore, the expression pattern of *Air-opnGqs* at the level of single photoreceptors also needs to be elucidated. Since *Air-opnGqs* are phylogenetically similar to the first reported scallop G_q-opsin in *Mizuhopecten yessoensis*, which is presumed to be co-expressed with G_q-protein in rhabdomeric photoreceptors of the proximal retina ("depolarizing layer") [48], we can predict that Air-OPNGqs will share a similar gross expression pattern. At a cellular level, it has been shown that more than one G_q-opsin can be expressed in a single photoreceptor cell [89–92] and this can lead to a broader spectral range for a given photoreceptor if opsins differ in λ_{max} values. Thus, to understand how spatial partitioning may have changed as gene copies diversified phenotypically in the scallop, future work will require the development of probes specific to each *Air-opnGq* gene or protein.

Spectral sensitivity may differ among the scallop G_q-opsin photopigments. We identified changes in amino acid sequence at seven sites that are predicted to influence spectral tuning of G_q-opsins [66]. The electrostatic contribution of individual residues at these sites has been modeled previously on Tpa-OPNGq1 [66, 75]. Among the scallop G_q-opsins, residues at position 92 had the most dissimilar biochemical properties (nonpolar aliphatic/hydrophobic in Air-OPNGq1 and Air-OPNGq2; aromatic in Air-OPNG3; positive polar in Air-OPNGq4). Position 306 is also of interest because there is a difference in charge and a presence/absence of a hydroxyl group. Air-OPNGq1 and Air-OPNGq4 have a polar, hydroxyl-bearing Thr306 while Air-OPNGq2 and Air-OPNGq3 contain a non-polar Ala306. Evidence from previous studies [93–95] suggests that shifts in λ_{max} values can be achieved via a change of charge (polar vs non-polar) or a gain/loss of a hydroxyl group that ultimately affects the electrostatic potential around

the protonated Schiff base [66]. Based on our results, we hypothesize that the λ_{max} may differ among some or all of the Air-OPNGqs. This hypothesis contradicts results from previous studies where only a single λ_{max} value was measured for depolarizing rhabdomeric photoreceptors [96, 97]. While some of the earliest work on spectral sensitivity of scallops was based on behavior trials, and was unable to test specific visual pigments, photoreceptor cells, or account for extra-ocular photoreception (e.g., [98]), more sophisticated methods have been employed to record membrane potential changes of individual photoreceptor cells (e.g., [97, 99, 100]). Most recently, microspectrophotometry has been used on dark-adapted scallop retinas to measure λ_{max} directly [96]. For rhabdomeric photoreceptors of *A. irradians*, both intracellular recordings [97] and microspectrophotometry results [96] recover a single spectral curve with a λ_{max} value of ~500 nm. Though, with the limited number of photoreceptor cells examined (N=4 versus N= 21 [96, 97]) and a 38- to 5815-fold higher expression level difference of *Air-opnGq2* to other *Air-opnGqs* (this study), it is unlikely that all four G_q-opsins were sampled. An alternative approach will be needed to determine if there are any differences in λ_{max} by targeting individual Air-OPNGqs. One approach would be to directly test λ_{max} of each Air-OPNGq photopigment *in vitro*, but the well-known technical challenges of expressing G_q-opsin proteins in transient heterologous systems will need to be overcome [101, 102] or stable transfection of cell lines [103] or animals [104] will need to be employed.

Conclusions

Gene duplication and subsequent functional divergence of opsins have played an important role in expanding photoreceptive capabilities of organisms by altering what wavelengths of light are preferentially absorbed by photoreceptors (spectral tuning). However, new opsin copies may also

acquire new or subdivide ancestral functions through changes to temporal, spatial or the level of gene expression. As the first molecular characterization of scallop G_q-opsins, our study highlights how opsin duplication and diversification may not only affect the evolution of the visual system, but also non-visual photoreception. Sequence variation among the scallop G_q-opsins suggests different biochemical properties of the proteins, which may translate into differences in light absorption and/or G protein affinity. Changes to spatial pattern and level of gene expression are illustrative of transitions between broad non-visual photoreception and eye-specific expression indicating neofunctionalization after opsin-duplication.

It is important to extend the taxonomic sampling of intraspecific opsin diversity in non-arthropod invertebrates in the future to understand diversification and plasticity of G_q-opsins. As such, molluscs are a rich system to study protein evolution, but have been underused due to a lack of basic information about their genic composition. Our work demonstrates the need for more studies looking at the visual evolution of molluscs to further their impact on the fields of molecular, sensory, and evolutionary biology.

Availability of data and material

Datasets supporting the results of this article are available from Genbank (KT426908 - KT426911).

Competing interests

The authors declare no competing interests.

Authors' contributions

JMS conceived the work; JMS, AJPK and ANP planned the research design; ANP collected samples; AJPK and ANP performed lab work; BSB, KS and DF performed protein modeling; ANP performed transcriptome assembly and phylogenetic analyses; all authors participated in manuscript preparation. All authors approve the final version of the manuscript.

Acknowledgments

We thank Eric Milbrant and the staff at the Sanibel-Captiva Conservation Foundation for organizing live scallop collection and providing housing and research resources, Brad Fleming for assisting in scallop collection and dissection, and Srihari Radhakrishnan for providing Trinity line code for transcriptome assembly. We thank Dan Speiser for commenting on an earlier version of the manuscript. This work was supported by the National Science Foundation (DEB 1118884 to JMS); the Carl A. and Grace A. Bailey Research Career Development Award (to JMS); the Iowa Science Foundation (ISF 11-13 to JMS and ANP); Sigma Xi (to ANP); and the Malacological Society of London (to ANP).

References

1. Briscoe AD, Macias-Muñoz A, Kozak KM, Walters JR, Yuan F, Jamie G a., Martin SH, Dasmahapatra KK, Ferguson LC, Mallet J, Jacquin-Joly E, Jiggins CD: **Female behaviour drives expression and evolution of gustatory receptors in butterflies**. *PLoS Genet* 2013, **9**.

2. Grus WE, Zhang J: **Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats**. *Gene* 2004, **340**:303–312.

3. Yokoyama S: **Molecular genetic basis of adaptive selection: examples from color vision in vertebrates**. *Annu Rev Genet* 1997, **31**:315–336.

4. Frentiu FD, Bernard GD, Sison-Mangus MP, Van Zandt Brower A, Briscoe AD: **Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies**. *Mol Biol Evol* 2007, **24**:2016–2028.

5. Dong D, Jones G, Zhang S: **Dynamic evolution of bitter taste receptor genes in vertebrates.** *BMC Evol Biol* 2009, **9**:12.

6. Niimura Y, Nei M: **Extensive gains and losses of olfactory receptor genes in mammalian evolution**. *PLoS One* 2007, **2**.

7. Ohno S: *Evolution by Gene Duplication*. Berlin: Springer-Verlag; 1970.

8. Walsh B: **Population-genetic models of the fates of duplicated genes**. *Genetica* 2003, **118**:279–294.

9. Hahn MW: **Distinguishing among evolutionary models for the maintenance of gene duplicates**. *J Hered* 2009, **100**:605–617.

10. Innan H, Kondrashov F: **The evolution of gene duplications: classifying and distinguishing between models.** *Nat Rev Genet* 2010, **11**:97–108.

11. Zhang J: **Evolution by gene duplication: An update**. *Trends Ecol Evol* 2003, **18**:292–298.

12. Force a, Force a, Lynch M, Lynch M, Postlethwait J, Postlethwait J: **Preservation of duplicate genes by subfunctionalization**. *Am Zool* 1999, **39**:0.

13. Spady TC, Parry JWL, Robinson PR, Hunt DM, Bowmaker JK, Carleton KL: **Evolution of the cichlid visual palette through ontogenetic subfunctionalization of the opsin gene arrays.** *Mol Biol Evol* 2006, **23**:1538–47.

14. Hittinger CT, Carroll SB: **Gene duplication and the adaptive evolution of a classic genetic switch.** *Nature* 2007, **449**:677–681.

15. Piatigorsky J, Wistow G: **The recruitment of crystallins: new functions precede gene duplication**. *Science (80-)* 1991, **252**:1078–1079.

16. Nathans J, Thomas D, Hogness DS: **Molecular genetics of human color vision: the genes encoding blue, green, and red pigments**. *Science (80-)* 1986, **232**:193–202.

17. Yokoyama S: **Molecular evolution of color vision in vertebrates.** *Gene* 2002, **300**:69–78.

18. Porter ML, Bok MJ, Robinson PR, Cronin TW: **Molecular diversity of visual pigments in Stomatopoda (Crustacea)**. *Vis Neurosci* 2009, **26**:255–265.

19. Briscoe AD: **Reconstructing the ancestral butterfly eye: focus on the opsins.** *J Exp Biol* 2008, **211**(Pt 11):1805–13.

20. O'Quin KE, Hofmann CM, Hofmann HA, Carleton KL: **Parallel evolution of opsin gene expression in African cichlid fishes.** *Mol Biol Evol* 2010, **27**:2839–2854.

21. Hofmann CM, Carleton KL: **Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish**. *Integr Comp Biol* 2009, **49**:630–643.

22. Futahashi R, Kawahara-Miki R, Kinoshita M, Yoshitake K, Yajima S, Arikawa K, Fukatsu T: **Extraordinary diversity of visual opsin genes in dragonflies**. *Proc Natl Acad Sci* 2015:201424670.

23. Palczewski K: **G protein-coupled receptor rhodopsin**. *Annu Rev Biochem* 2006, **75**:743–767.

24. Yarfitz S, Hurley JB: **Transduction mechanisms of vertebrate and invertebrate photoreceptors**. *J Biol Chem* 1994, **269**:14329–14332.

25. Marin EP: **The amino terminus of the fourth cytoplasmic loop of rhodopsin modulates rhodopsin-transducin interaction**. *J Biol Chem* 2000, **275**:1930–1936.

26. Plachetzki DC, Degnan BM, Oakley TH: **The origins of novel protein interactions during animal opsin evolution**. *PLoS One* 2007, **2**: e1054.

27. Porter ML, Blasic JR, Bok MJ, Cameron EG, Pringle T, Cronin TW, Robinson PR: **Shedding new light on opsin evolution.** *Proc Biol Sci* 2012, **279**:3–14.

28. Feuda R, Rota-Stebelli O, Oakley TH, Pisani D: **The comb jelly opsins and the origins of animal phototransduction.** *Genome Biol Evol* 2014, **6**:1964–1971.

29. Cronin TW, Porter ML: **The evolution of invertebrate photopigments and photoreceptors**. In *Evol Vis Non-visual Pigment*. Edited by Hunt DM, Hankins MW, Collin SP, Marshall NJ. New York: Springer International Publishing; 2014:105–135.

30. Fuller RC, Carleton KL, Fadool JM, Spady TC, Travis J: **Genetic and environmental variation in the visual properties of bluefin killifish, Lucania goodei.** *J Evol Biol* 2005, **18**:516–23.

31. Rennison DJ, Owens GL, Taylor JS: **Opsin gene duplication and divergence in ray-finned fish.** *Mol Phylogenet Evol* 2012, **62**:986–1008.

32. Dulai KS, von Dornum M, Mollon JD, Hunt DM: **The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates.** *Genome Res* 1999, **9**:629–38.

33. Briscoe AD: **Functional diversification of lepidopteran opsins following gene duplication**. *Mol Biol Evol* 2001, **18**:2270–2279.

34. Koyanagi M, Nagata T, Katoh K, Yamashita S, Tokunaga F: **Molecular evolution of arthropod color vision deduced from multiple opsin genes of jumping spiders**. *J Mol Evol* 2008, **66**:130–137.

35. Chinen A, Hamaoka T, Yamada Y, Kawamura S: **Gene duplication and spectral diversification of cone visual pigments of zebrafish**. *Genetics* 2003, **163**:663–675.

36. Carulli JP, Chen DM, Stark WS, Hartl DL: **Phylogeny and physiology of Drosophila opsins**. *J Mol Evol* 1994, **38**:250–262.

37. Porter ML, Cronin TW, McClellan DA, Crandall KA: **Molecular characterization of crustacean visual pigments and the evolution of pancrustacean opsins**. *Mol Biol Evol* 2007, **24**:253–268.

38. Wang D, Oakley T, Mower J, Shimmin LC, Yim S, Honeycutt RL, Tsao H, Li WH: **Molecular Evolution of Bat Color Vision Genes**. *Mol Biol Evol* 2004, **21**:295–302.

39. Cortesi F, Musilová Z, Stieb SM, Hart NS, Siebeck UE, Malmstrøm M, Tørresen OK, Jentoft S, Cheney KL, Marshall NJ, Carleton KL, Salzburger W: **Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes**. *Proc Natl Acad Sci* 2015, **112**:1493–1498.

40. Yokoyama S: **Gene duplications and evolution of the short wavelength-sensitive visual pigments in vertebrates.** *Mol Biol Evol* 1994, **11**:32–39.

41. Oakley TH, Huber DR: **Differential expression of duplicated opsin genes in two eye types of ostracod crustaceans**. *J Mol Evol* 2004, **58**:1–11.

42. Spaethe J, Briscoe AD: **Early duplication and function diversification of the opsin gene family in insects**. *Mol Biol Evol* 2004, **21**:1583–1594.

43. Henze MJ, Dannenhauer K, Kohler M, Labhart T, Gesemann M: **Opsin evolution and expression in Arthropod compound Eyes and Ocelli: Insights from the cricket Gryllus bimaculatus**. *BMC Evol Biol* 2012, **12**:163.

44. Pollock J, Benzer S: **Transcript localization of four opsin genes in the three visual organs of Drosophila; RH2 is ocellus specific**. *Nature* 1988.

45. Tong D, Rozas NS, Oakley TH, Mitchell J, Colley NJ, McFall-Ngai MJ: **Evidence for light perception in a bioluminescent organ**. *Proc Natl Acad Sci USA* 2009, **106**:9836–9841.

46. Frank TM, Porter M, Cronin TW: **Spectral sensitivity, visual pigments and screening pigments in two life history stages of the ontogenetic migrator Gnathophausia ingens**. *J Mar Biol Assoc United Kingdom* 2009, **89**:119–129.

47. Rivera AS, Pankey MS, Plachetzki DC, Villacorta C, Syme AE, Serb JM, Omilian AR, Oakley TH: **Gene duplication and the orgins of morphological complexity in pancrustacean eyes, a genomic approach**. *BMC Evol Biol* 2010, **10**:123.

48. Kojima D, Terakita A, Ishikawa T, Tsukahara Y, Maeda A, Shichida Y: **A novel Go-mediated phototransduction casade in scallop visual cells**. *J Biol Chem* 1997, **272**:22979–22982.

49. Serb JM, Porath-Krause AJ, Pairett AN: **Uncovering a gene duplication of the photoreceptive protein, opsin, in scallops (Bivalvia: Pectinidae).** *Integr Comp Biol* 2013, **53**:68–77.

50. Kimura M: *The Neutral Theory of Molecular Evolution*. Cambridge; 1983.

51. Murakami M, Kouyama T: **Crystal structure of squid rhodopsin**. *Nature* 2008, **453**:363–367.

52. Shimamura T, Hiraki K, Takahashi N, Hori T, Ago H, Masuda K, Takio K, Ishiguro M, Miyano M: **Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region.** *J Biol Chem* 2008, **283**:17753–6.

53. Dalal JS, Jinks RN, Cacciatore C, Greenberg RM, Battelle B-A: **Limulus opsins: diurnal regulation of expression**. *Vis Neurosci* 2003, **20**:523–534.

54. Halstenberg S, Lindgre K, Samagh S, Nadal-Vicens M, Balt S, Fernald RD: **Diurnal rhythm of cone opsin expression in the teleost fish Haplochromis burtoni**. *Vis Neurosci* 2005, **22**:135–141.

55. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A: **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.** *Nat Protoc* 2013, **8**:1494–512.

56. Pairett AN, Serb JM: **De novo assembly and characterization of two transcriptomes reveal multiple light-mediated functions in the scallop eye (Bivalvia: Pectinidae).** *PLoS One* 2013, **8**:e69852.

57. Katoh K, Kuma K, Toh H, Miyata T: **MAFFT version 5: improvement in accuracy of multiple sequence alignment**. *Nucleic Acids Res* 2005, **33**:511–518.

58. Abascal F, Zardoya R, Posada D: **ProtTest: Selection of best-fit models of protein evolution**. *Bioinformatics* 2005, **21**:2104–2105.

59. Le S, Gascuel O: **An improved general amino acid replacement matrix**. *Mol Biol Evol* 2008, **25**:1307–1320.

60. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics* 2014, **30**:1412–1422.

61. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny**. *Bioinformatics* 2001, **17**:754–755.

62. Miller M, Pfeiffer W, Schwartz T: **Creating the CIPRES Science Gateway for inference of large phylogenetic trees**. In *Proc Gatew Comput Environ Work*; 2010:1–8.

63. Lonnig W-E, Saedler H: **Chromosome rearrangements and transposable elements.** *Annu Rev Genet* 2002, **36**:389–410.

64. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein search programs**. *Nucleic Acids Res* 1997, **25**:3389–3402.

65. Altschul S, Wooten J, Gertz E, Agarwala R, Morgulis A, Schaffer A, Yu Y-K: **Protein database searches using compositionally adjusted substitution matrices**. *FEBS* 2005, **272**:5101–5109.

66. Sekharan S, Wei JN, Batista VS: **The active site of melanopsin: the biological clock photoreceptor.** *J Am Chem Soc* 2012, **134**:19536–9.

67. Zhang Y: **I-TASSER server for protein 3D structure prediction**. *BMC Bioinformatics* 2008, **9**:40.

68. Roy A, Kucikural A, Zhang Y: **I-TASSER: a unified platform for automated protein structure and function prediction**. *Nat Protoc* 2010, **5**:725–738.

69. Laskowski R, MacArthur M, Moss D, Thornton J: **PROCHECK: a program to check the stereochemical quality of protein structures**. *J Appl Crystallogr* 1993, **26**:283–291.

70. Laskowski R, Hutchinson E, Michie A, Wallace A, Jones M, Thornton J: **PDBsum: a Web-based database of summaries and analyses of all PDB structures**. *Trends Biochem Sci* 1997, **22**:488–490.

71. Langmead B, Trapnell C, Pop M, Salzberg S: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome**. *Genome Biol* 2009, **10**:R25.

72. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12**:323.

73. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, et al.: **The oyster genome reveals stress adaptation and complexity of shell formation.** *Nature* 2012, **490**:49–54.

74. Rosenbaum DM, Rasmussen SGF, Kobilka BK: **The structure and function of G-protein-coupled receptors.** *Nature* 2009, **459**:356–363.

75. Sekharan S, Altun A, Morokuma K: **Photochemistry of visual pigment in a G(q) protein-coupled receptor (GPCR)--insights from structural and spectral tuning studies on squid rhodopsin.** *Chem - a Eur J* 2010, **16**:1744–9.

76. Vogel R, Mahalingam M, Lüdeke S, Huber T, Siebert F, Sakmar TP: **Functional Role of the "Ionic Lock"-An Interhelical Hydrogen-Bond Network in Family A Heptahelical Receptors**. *J Mol Biol* 2008, **380**:648–655.

77. Plachetzki DC, Oakley TH: **Key transitions during the evolution of animal phototransduction: novelty, "tree-thinking," co-option, and co-duplication**. *Integr Comp Biol* 2007, **47**:759–769.

78. Briscoe AD, Chittka L: **The evolution of color vision in insects**. *Annu Rev Entomol* 2001, **46**:471–510.

79. Hart N, Hunt D: **Avian Visual Pigments: Characteristics, Spectral Tuning, and Evolution**. *Am Nat* 2007, **169**:S7–S26.

80. Bowmaker JK: **Evolution of vertebrate visual pigments**. *Vision Res* 2008, **48**:2022–2041.

81. O'Quin KE, Smith AR, Sharma A, Carleton KL: **New evidence for the role of heterochrony in the repeated evolution of cichlid opsin expression**. *Evol Dev* 2011, **13**:193–203.

82. Wang Y, Guo X: **Chromosomal rearrangement in pectinidae revealed by rRNA loci and implications for bivalve evolution.** *Biol Bull* 2004, **207**:247–56.

83. Zhang L, Bao Z, Wang S, Huang X, Hu J: **Chromosome rearrangements in Pectinidae (Bivalvia: Pteriomorphia) implied based on chromosomal localization of histone H3 gene in four scallops.** *Genetica* 2007, **130**:193–8.

84. Wilkens LA: **Primary inhibition by light : A unique property of bivalve photoreceptors**. *Am Malacol Bull* 2008, **26**:101–109.

85. Morton B: **The evolution of eyes in the Bivalvia**. *Oceanogr Mar Biol an Annu Rev* 2001, **39**:165–205.

86. Speiser DI, Johnsen S: **Comparative morphology of the concave mirror eyes of scallops (Pectinoidea)**. *Am Malacol Bull* 2008, **26**:27–33.

87. Gutsell JS: **Natural history of the bay scallop**. *Bull Bur Fish* 1930, **46**:569–632.

88. Wilkens LA: **Neurobiology and behavior of the scallop**. In *Scallops Biol Ecol Aquac*. Edited by Shumway SE, Parsons GJ. Elsevier; 2006:317–356.

89. Mazzoni EO, Celik A, Wernet MF, Vasiliauskas D, Johnston RJ, Cook T a., Pichaud F, Desplan C: **Iroquois complex genes induce co-expression of rhodopsins in Drosophila**. *PLoS Biol* 2008, **6**:825–835.

90. Arikawa K, Mizuno S, Kinoshita M, Stavenga DG: **Coexpression of two visual pigments in a photoreceptor causes an abnormally broad spectral sensitivity in the eye of the butterfly Papilio xuthus.** *J Neurosci* 2003, **23**:4527–4532.

91. Katti C, Kempler K, Porter ML, Legg A, Gonzalez R, Garcia-Rivera E, Dugger D, Battelle B: **Opsin co-expression in Limulus photoreceptors: differential regulation by light and a circadian clock.** *J Exp Biol* 2010, **213**(Pt 15):2589–601.

92. Hu X, Leming MT, Whaley M a, O'Tousa JE: **Rhodopsin coexpression in UV photoreceptors of Aedes aegypti and Anopheles gambiae mosquitoes**. *J Exp Biol* 2014, **217**:1003–1008.

93. Asenjo AB, Rim J, Oprian DD: **Molecular determinants of human red/green color discrimination**. *Neuron* 1994, **12**:1131–1138.

94. Yokoyama S, Tada T, Zhang H, Britt L: **Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates**. *Proc Natl Acad Sci USA* 2008, **105**:13480–13485.

95. Hauser FE, van Hazel I, Chang BSW: **Spectral tuning in vertebrate short wavelength-sensitive 1 (SWS1) visual pigments: Can wavelength sensitivity be inferred from sequence data?** *J Exp Zool Part B Mol Dev Evol* 2014, **3228**:529–539.

96. Speiser DI, Loew ER, Johnsen S: **Spectral sensitivity of the concave mirror eyes of scallops: potential influences of habitat, self-screening and longitudinal chromatic aberration.** *J Exp Biol* 2011, **214**(Pt 3):422–31.

97. McReynolds JS, Gorman ALF: **Membrane conductances and spectral sensitivities of Pecten photoreceptors**. *J Gen Physiol* 1970, **56**:392–406.

98. Cronly-Dillon JR, Cronly-Dillion JR: **Spectral sensitivity of the scallop Pecten maximus**. *Science (80-)* 1966, **151**:345–346.

99. Cornwall MC, Gorman ALF: **The cation selectivity and voltage dependence of the light-activated potassium conductance in scallop distal photoreceptor**. *J Physiol* 1983, **340**:287–305.

100. Gomez MP, Nasi E: **The light-sensitive conductance of hyperpolarizing invertebrate photoreceptors: a patch-clamp study**. *J Gen Physiol* 1994, **103**:939–956.

101. Terakita A, Tsukamoto H, Koyanagi M, Sugahara M, Yamashita T, Shichida Y: **Expression and comparative characterization of Gq-coupled invertebrate visual pigments and melanopsin**. *J Neurochem* 2008, **105**:883–890.

102. Matsuyama T, Yamashita T, Imamoto Y, Shichida Y: **Photochemical properties of mammalian melanopsin.** *Biochemistry* 2012, **51**:5454–62.

103. Frentiu FD, Yuan F, Savage WK, Bernard GD, Mullen SP, Briscoe a. D: **Opsin Clines in Butterflies Suggest Novel Roles for Insect Photopigments**. *Mol Biol Evol* 2014, **32**:368–379.

104. Knox BE, Salcedo E, Mathiesz K, Schaefer J, Chou W-H, Chadwell L V, Smith WC, Britt SG, Barlow RB: **Heterologous expression of limulus rhodopsin.** *J Biol Chem* 2003, **278**:40493–502.

105. Waterhouse A, Procter J, Martin D, Clamp M, Barton GJ: **Jalview Version 2 - a multiple sequence alignment editor and analysis workbench**. *Bioinformatics* 2009, **25**:1189–1191.

Table 2-1 Percent similarity (below diagonal) and RMSD (above diagonal) of scallop (Air) and squid (Tpa) proteins.

	Air-OPNGq1	Air-OPNGq2	Air-OPNGq3	Air-OPNGq4	Tpa-OPNGq1
Air-OPNGq1	-	0.378[b]	**0.354**	0.489	0.589
Air-OPNGq2	74.7 (78.9)[a]	-	0.408	0.603	0.503
Air-OPNGq3	74.7 (77.9)	**80.9 (92.6)**[c]	-	**0.699**	0.601
Air-OPNGq4	72.9 (76.9)	76.9 (85.6)	74.6 (88.1)	-	0.549
Tpa-OPNGq1	**71.0 (72.4)**	73.4 (73.8)	75.2 (75.8)	73.3 (74.7)	-

[a] Percent similarity of amino acid sequence alignments from first methionine to stop codons; values in parentheses are percent identity from Helix 1 through Helix 7

[b] Atomic values in angstroms, where the lower the RMSD value, the higher is the similarity between structures

[c] Numbers in bold indicate minimum and maximum values

Table 2-2 Sequence and structural motifs in scallop (Air) and squid (Tpa) G_q-opsins

Motifs	Air-OPNGq1 493 aa	Air-OPNGq2 456 aa	Air-OPNGq3 519 aa	Air-OPNGq4 481 aa	Tpa-OPNGq1 448 aa
LxxxD TMII (pos 76-80)	LAVSD	LALSD	LALSD	LALSD	LAFSD
Disulfide bond	C108, C186	C108, C186	C108, C186	C108, C186	C108, C186
Hydrogen bond with Schiff base	N87, Y111	N87, Y111	N87, Y111	N87, Y111	N87, Y111
E/DRY TMIII (pos 132-134)	DRY	DRC	DRF	DRF	DRY
Counterion	E180	E180	E180	E180	E180
LAK TMVII (pos 305-307)	LAK	LAK	LAK	LAK	FAK
NPxxY TMVII (pos 311-315)	NPIIY	NPIVY	NPIVY	NPLVY	NPMIY
G-protein binding (pos 319-321)	HPK	HPK	HPK	HPR	HPK

Note – The amino acid numbering system follows the amino acid position (pos) of squid rhodopsin.

Table 2-S1 Primers used to amplify scallop G_q-opsins and intergenic region between *Air-opnGq3* and *Air-opnGq4*.

Gene/ fragment	Primer Name	Sequence (5'-3')	Material	Product Length (nt)
Air-opnGq1	212F	CGCCTTATCACTCCGCAC	cDNA	1630
	1841R	CCTCAGTCATATCCAAGATGCC		
Air-opnGq1	212F	CGCCTTATCACTCCGCAC	gDNA	1009
	1100R	GATGGCTGAGAGCATATACCACTGG		
Air-opnGq2	38F	GCTACCACCCAGGGTACTCC	cDNA	1585
	1552R	CAGCAAACACGTGACTTCCATC		
Air-opnGq2	38F	GCTACCACCCAGGGTACTCC	gDNA	1585
	1552R	CAGCAAACACGTGACTTCCATC		
Air-opnGq3	807F	CCGTTGGGGTTTATCATCTC	cDNA	1741
	2548R	GACAAACTGGAAATGAACTC		
Air-opnGq3	807F	CCGTTGGGGTTTATCATCTC	gDNA	1825
	2632R	CCTGGATATGTCGATGC		
Air-opnGq4	66F	CGTGCCCAACTTCATCATC	cDNA	1609
	97R	CAAAAGTGTGCGTTCTTCTG		
Air-opnGq4	66F	CGTGCCCAACTTCATCATC	gDNA	1609
	97R	CAAAAGTGTGCGTTCTTCTG		
Intergenic region	PR-4-Fw-Air-OPNGq3	TAATATAGCCTTGCCTGCCTCACTTGCC	cDNA	1752
	PR-5-Rv_Air-OPNGq4	GGAACAATCTCTGTCACAGATTCCTCGG		
Intergenic region	PR-4-Fw-Air-OPNGq3	TAATATAGCCTTGCCTGCCTCACTTGCC	gDNA	1752
	PR-5-Rv_Air-OPNGq4	GGAACAATCTCTGTCACAGATTCCTCGG		

Table 2-S2 G_q-opsin sequences included in the phylogenetic analysis.

Identity	Phylum	Genus	species	common name	Genbank accession# or source
Al.subulata	Mollusca	*Alloteuthis*	*subulata*	squid	Z49108
Ap.californica 1	Mollusca	*Aplysia*	*californica*	sea hare	AASC01108363*
Ap.californica 2	Mollusca	*Aplysia*	*californica*	sea hare	AASC02005512*
Air-OPNGo1	Mollusca	*Argopecten*	*irradians*	bay scallop	TBD
Air-OPNGq1	Mollusca	*Argopecten*	*irradians*	bay scallop	KT426908
Air-OPSGq2	Mollusca	*Argopecten*	*irradians*	bay scallop	KT426909
Air-OPNGq3	Mollusca	*Argopecten*	*irradians*	bay scallop	KT426910
Air-OPNGq4	Mollusca	*Argopecten*	*irradians*	bay scallop	KT426911
Bi.glabrata2	Mollusca	*Biomphalaria*	*glabrata*	ram's horn snail	Genome, Dejong et al. 2004
Bi.glabrata1	Mollusca	*Biomphalaria*	*glabrata*	ram's horn snail	Genome, Dejong et al. 2004
Cgi-OPNGq2B	Mollusca	*Crassostrea*	*gigas*	pacific oyster	Genome, Zhang et al. 2012
Cgi-OPNGq1	Mollusca	*Crassostrea*	*gigas*	pacific oyster	Genome, Zhang et al. 2012
Cgi-OPNGq2A	Mollusca	*Crassostrea*	*gigas*	pacific oyster	Genome, Zhang et al. 2012
En.dolfeini	Mollusca	*Enteroctopus*	*dofleini*	octopus	CAA30644.1
Eu.scolopes	Mollusca	*Euprymna*	*scolopes*	ceph	ACB05673.1
Lo.forbesi	Mollusca	*Loligo*	*forbesi*	squid	CAA40108.1
Lo.pealei	Mollusca	*Loligo*	*pealei*	squid	AY450853
L.gigantea1	Mollusca	*Lottia*	*gigantea*	limpet	FC774055
L.gigantea2	Mollusca	*Lottia*	*gigantea*	limpet	Genome*
Mye-OPNGq1	Mollusca	*Mizuhopecten*	*yessoensis*	scallop	AB006454
Pi.fucata1	Mollusca	*Pinctata*	*fucata*	pearl oyster	Genome, Takeuchi et al 2012
Pi.fucata2	Mollusca	*Pinctata*	*fucata*	pearl oyster	Genome, Takeuchi et al 2012
Pma-OPNGq2	Mollusca	*Placopecten*	*magellanicus*	sea scallop	Pairett & Serb 2013
Pma-OPNGq3	Mollusca	*Placopecten*	*magellanicus*	sea scallop	Pairett & Serb 2013
Se.officinalis	Mollusca	*Sepia*	*officinalis*	cuttlefish	AF000947
Tpa-OPNGq1	Mollusca	*Todarodes*	*pacificus*	squid	X70498
Ca.capitata	Annelida	*Capitella*	*capitata*	polychaete worm	Genome*
He.robusta1	Annelida	*Helobdella*	*robusta*	leech	Genome*

Table 2-S2 (continued)

Identity	Phylum	Genus	species	common name	Genbank accession# or source
He.robusta2	Annelida	*Helobdella*	*robusta*	leech	Genome, scaffold_391*
P.dumerilii	Annelida	*Platynereis*	*dumerilii*	ragworm	AJ316544
An.gambiae	Arthropoda	*Anopheles*	*gambiae*	mosquito	CAA76825.1
An.gambiaeUV5	Arthropoda	*Anopheles*	*gambiae*	mosquito	XP_001688790
An.gambiaeUV7	Arthropoda	*Anopheles*	*gambiae*	mosquito	XP_308329
An.gambiaeUVB	Arthropoda	*Anopheles*	*gambiae*	mosquito	XP_319247.1
Ap.melliferaA	Arthropoda	*Apis*	*mellifera*	bee	NM_001077825
Ap.melliferaB	Arthropoda	*Apis*	*mellifera*	bee	U26026
Ap.melliferaUV5	Arthropoda	*Apis*	*mellifera*	bee	AAC13418
Ap.melliferaUVB	Arthropoda	*Apis*	*mellifera*	bee	AF004168
B.anynana	Arthropoda	*Bicyclus*	*anynana*	butterfly	157502893
B.anynanaUV	Arthropoda	*Bicyclus*	*anynana*	butterfly	AAL91507.1
B.anynanaUVB	Arthropoda	*Bicyclus*	*anynana*	butterfly	AAY16527.1
D.pulexBCR	Arthropoda	*Daphnia*	*pulex*	water flea	GL732562.1
D.pulex	Arthropoda	*Daphnia*	*pulex*	water flea	Genome*
D.pulexUV5a	Arthropoda	*Daphnia*	*pulex*	water flea	EFX75461
D.pulexUV5b	Arthropoda	*Daphnia*	*pulex*	water flea	EFX81332
Dr.melanogast1	Arthropoda	*Drosophila*	*melanogaster*	fruitfly	NP_524407.1
Dr.melanogast2	Arthropoda	*Drosophila*	*melanogaster*	fruitfly	AAA28734.1
Dr.melanogast6	Arthropoda	*Drosophila*	*melanogaster*	fruitfly	CAB06821.1
Dr.melanogastUV3	Arthropoda	*Drosophila*	*melanogaster*	fruitfly	AAA28854.1
Dr.melanogastUV4	Arthropoda	*Drosophila*	*melanogaster*	fruitfly	NP_476701.1
Dr.melanogastUV5	Arthropoda	*Drosophila*	*melanogaster*	fruitfly	AAC47426.1
Dr.melanogastUV7	Arthropoda	*Drosophila*	*melanogaster*	fruitfly	NP_524035
Ha.adansoni1	Arthropoda	*Hasarius*	*adansoni*	jumping spider	BAG14330.1
Ha.adansoni2	Arthropoda	*Hasarius*	*adansoni*	jumping spider	BAG14331.1
Ha.adansoniUV5	Arthropoda	*Hasarius*	*adansoni*	jumping spider	BAG14332.1

Table 2-S2 (continued)

Identity	Phylum	Genus	species	common name	Genbank accession# or source
Ix.scapularis	Arthropoda	*Ixodes*	*scapularis*	tick	XM_002408275.1
Ix.scapularisUV7	Arthropoda	*Ixodes*	*scapularis*	tick	Genome*
Li.polyphemusBCR	Arthropoda	*Limulus*	*polyphemus*	horseshoe crab	ACO05013
Li.polyphemus	Arthropoda	*Limulus*	*polyphemus*	horseshoe crab	AAA02498.1
Ne.oerstedii2	Arthropoda	*Neogonodactylus*	*oerstedii*	stomatopod	DQ646870
Ne.oerstedii3	Arthropoda	*Neogonodactylus*	*oerstedii*	stomatopod	DQ646871
Ne.oerstedii1	Arthropoda	*Neogonodactylus*	*oerstedii*	stomatopod	DQ646869
Pl.paykulli1	Arthropoda	*Plexippus*	*paykulli*	jumping spider	BAG14333.1
Pl.paykulli2	Arthropoda	*Plexippus*	*paykulli*	jumping spider	BAG14334.1
Pl.paykulliUV5	Arthropoda	*Plexippus*	*paykulli*	jumping spider	BAG14335
Tr.castaneum	Arthropoda	*Tribolium*	*castaneum*	flour beetle	ABA00706.1
Tr.castaneumUV5	Arthropoda	*Tribolium*	*castaneum*	flour beetle	ABW06837.1
Br.belcheri6	Chordata	*Branchiostoma*	*belcheri*	amphioxus	AB050611
Br.floridae6	Chordata	*Branchiostoma*	*floridae*	amphioxus	XP_002586119.1
Br.belcheri	Chordata	*Branchiostoma*	*belcheri*	amphioxus	AB205400
Br.floridae	Chordata	*Branchiostoma*	*floridae*	amphioxus	Genome*
Ci.intestinalis	Chordata	*Ciona*	*intestinalis*	tunicate	AABS01000008.1
Ci.savignyi	Chordata	*Ciona*	*savignyi*	tunicate	Genome*
Da.rerio1A	Chordata	*Danio*	*rerio*	zebrafish	Genome*
Da.rerio1B	Chordata	*Danio*	*rerio*	zebrafish	Genome*
Da.rerio2	Chordata	*Danio*	*rerio*	zebrafish	Genome*
Da.rerioGt.MWS	Chordata	*Danio*	*rerio*	zebrafish	NP_571250.1
Da.rerioGt.PAR	Chordata	*Danio*	*rerio*	zebrafish	XP_003201482
Da.rerioGt.PPIN	Chordata	*Danio*	*rerio*	zebrafish	NP_001005312.1
Da.rerioGt.Rho1	Chordata	*Danio*	*rerio*	zebrafish	BC164171.1
Da.rerioGt.Rho2	Chordata	*Danio*	*rerio*	zebrafish	NM_131254
Da.rerioGt.SWS	Chordata	*Danio*	*rerio*	zebrafish	NP_571394.1

Table 2-S2 (continued)

Identity	Phylum	Genus	species	common name	Genbank accession# or source
Da.rerioGt.TMT	Chordata	*Danio*	*rerio*	zebrafish	NP_001112371.1
Da.rerioGt.VA	Chordata	*Danio*	*rerio*	zebrafish	NM_131586
Ga.gallus1	Chordata	*Gallus*	*gallus*	chicken	NP_001038118.1
Ga.gallus2	Chordata	*Gallus*	*gallus*	chicken	AY882944
Mu.musculus	Chordata	*Mus*	*musculus*	mouse	AF147789
Xe.laevis1	Chordata	*Xenopus*	*laevis*	frog	ABD37674.1
Xe.laevis2	Chordata	*Xenopus*	*laevis*	frog	Genome*
St.purpuratus	Echino-dermata	*Strongylo-centrotus*	*purpuratus*	sea urchin	XR_026330*
Du.japonica	Platy-helminthes	*Dugesia*	*japonica*	flatworm	CAD13146
Gi.tigrina	Platy-helminthes	*Girardia*	*tigrina*	flatworm	CAB89516
Sc.mansoni1	Platy-helminthes	*Schistosoma*	*mansoni*	trematode worm	AF155134
Sc.mansoni2	Platy-helminthes	*Schistosoma*	*mansoni*	trematode worm	CD096414
Sc.mansoni3	Platy-helminthes	*Schistosoma*	*mansoni*	trematode worm	Smp_180030
Sc.meditrranea	Platy-helminthes	*Schmidtea*	*mediterranea*	planaria	AF112361

Asterisks represent sequences obtained through Porter *et al*. [27]. For additional information regarding sequence acquisition not available on Genbank, see supplementary material in Porter *et al*. [27].

Table 2-S3 Ramachandran plot values and C-scores for top G_q-opsin models.

For each Air-OPNGq, the top five models reported by I-TASSER were analyzed for their quality using PROCHECK and the C-score. All the reported models have > 90% of their residues in allowed regions of the Ramachandran plot, indicating a good quality model. The C-scores for the best models was in the range of -3 to -2. While these values are lower than the suggested cutoff of -1.5, this is not unexpected for GPCRs because there are relatively few solved GPCR protein structures and GPCRs often show high sequence diversity. The best model for each Air-OPNGq (highlighted) was selected as the structure having the highest C-score and highest percentage of residues in allowed regions of the Ramachandran plot.

	Model number	% residues in most favored regions	% residues in additionally allowed regions	% residues in generously allowed regions	% residues in disallowed regions	C-scores
Air-OPNGq1	**1**	**85.4**	**10.9**	**2.5**	**1.2**	**-1.95**
	2	87.3	8.6	2.3	1.9	-2.05
	3	83.6	12.7	2.1	1.6	-2.53
	4	84.7	10.6	3.7	0.9	-3.06
	5	85.4	10.2	2.1	2.3	-2.53
Air-OPNGq2	**1**	**85.6**	**10.6**	**1.8**	**2.0**	**-0.85**
	2	87.4	7.6	2.3	2.8	-2.01
	3	85.9	7.1	4.8	2.3	-2.22
	4	87.9	8.8	1.5	1.8	-1.35
	5	87.4	7.8	3.3	1.5	-2.00
Air-OPNGq3	**1**	**83.6**	**10.7**	**3.3**	**2.4**	**-2.17**
	2	86.0	9.2	2.6	2.2	-2.33
	3	88.8	8.3	1.5	1.3	-2.44
	4	84.0	11.6	3.5	0.9	-2.56
	5	86.0	7.9	3.5	2.6	-2.70
Air-OPNGq4	**1**	**87.2**	**8.5**	**2.4**	**1.9**	**-1.98**
	2	86.7	7.8	3.1	2.4	-2.25
	3	84.4	12.6	2.6	0.5	-2.42
	4	83.9	11.4	2.1	2.6	-2.37
	5	88.2	6.9	4.0	0.9	-2.34
Tpa-OPNGq1		70.4	27.1	2.1	0.3	

Table 2-S3 (continued)

Note – For each Air-OPNGq, the top five models reported by I-TASSER were analyzed for their quality using PROCHECK and the C-score. All the reported models have > 90% of their residues in allowed regions of the Ramachandran plot, indicating a good quality model. The C-scores for the best models was in the range of -3 to -2. While these values are lower than the suggested cutoff of -1.5, this is not unexpected for GPCRs because there are relatively few solved GPCR protein structures and GPCRs often show high sequence diversity. The best model for each Air-OPNGq (highlighted) was selected as the structure having the highest C-score and highest percentage of residues in allowed regions of the Ramachandran plot.

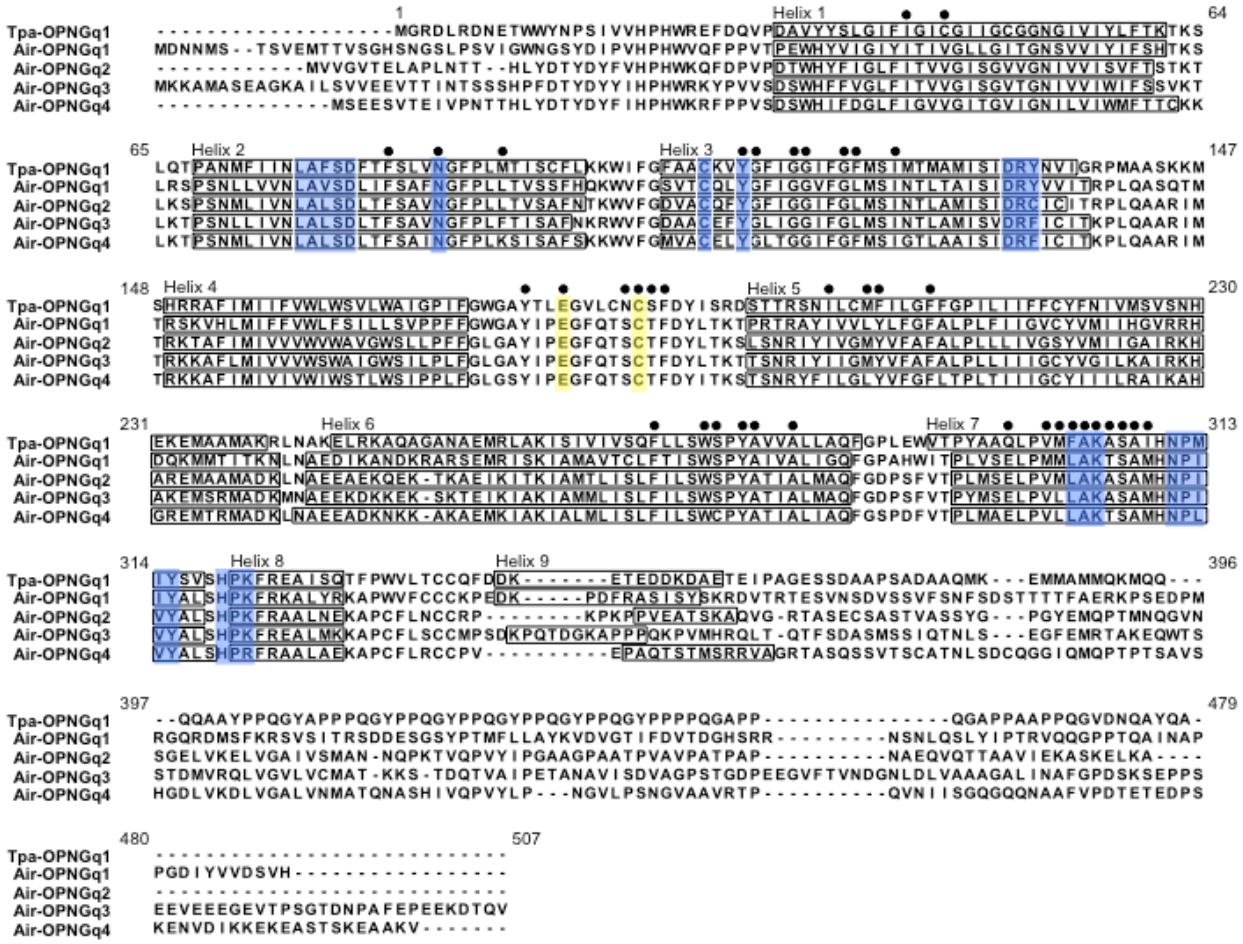


Figure 2-1 Amino acid alignment of G_q-opsins from scallop (Air-OPNGq1-OPNGq4) and squid, *Todarodes pacificus* (Tpa-OPNGq1). The alpha-helix domains are based on protein structure homology modeling (this study) or have been adapted from Shimamura *et al*. [52]. Sequence motifs described in Table 2-1 are in yellow; residues important for structural confirmation are in blue. Numbering of amino acid positions begins with the start codon (Met) of Tpa-OPNGq1.

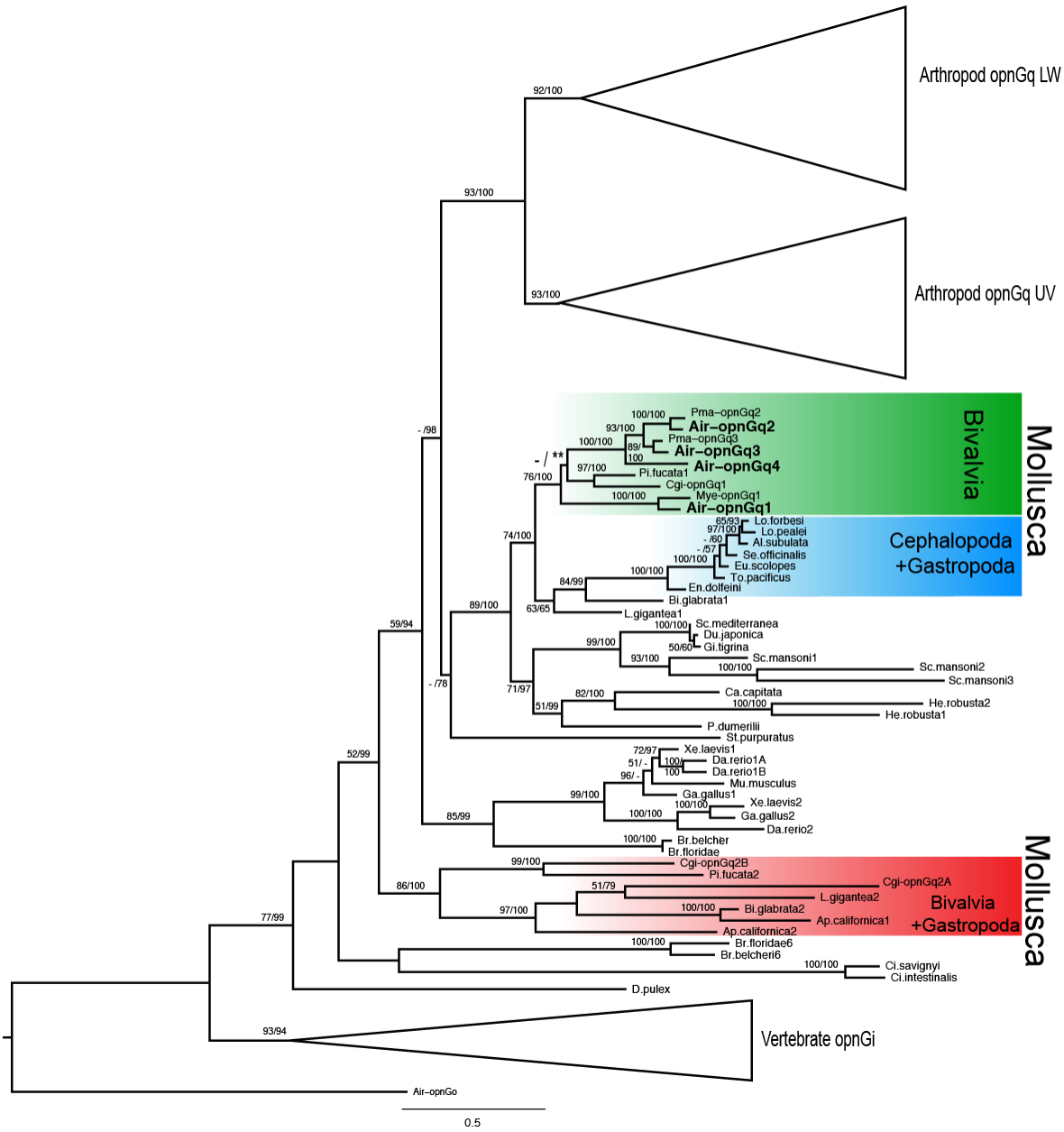


Figure 2-2 Maximum likelihood (ML) topology of G_q-opsins. The phylogenetic tree is based on aligned amino acid sequences with scallop G_o-opsin as the outgroup. Support values (>50%) of nodes were generated by 1000 bootstrap replicates in RAxML. Support values after the '/' are posterior probabilities from a Bayesian phylogenetic analysis (BI). Support values <50% are indicated by a '-'. The single difference between the ML and BI topologies occurs within the

bivalve G_q-opsin clade (green) and is highlighted with an asterisk '*'. *Argopecten irradians* G_q-opsins (Air-OPNGqs) from this study are in bold. Two molluscan G_q-opsin clades were recovered, but only one clade (green) contained scallop G_q-opsins from *Argopecten irradians* (Air-OPNGqs), *Mizuhopecten yessoensis* (Mye-OPNGqs), or *Placopecten magellanicus* (Pma-OPNGqs). Two large clades of arthropod opsins that represent UV and long-wavelength (LW) opsins and a vertebrate G_i–opsin clade were collapsed for space.

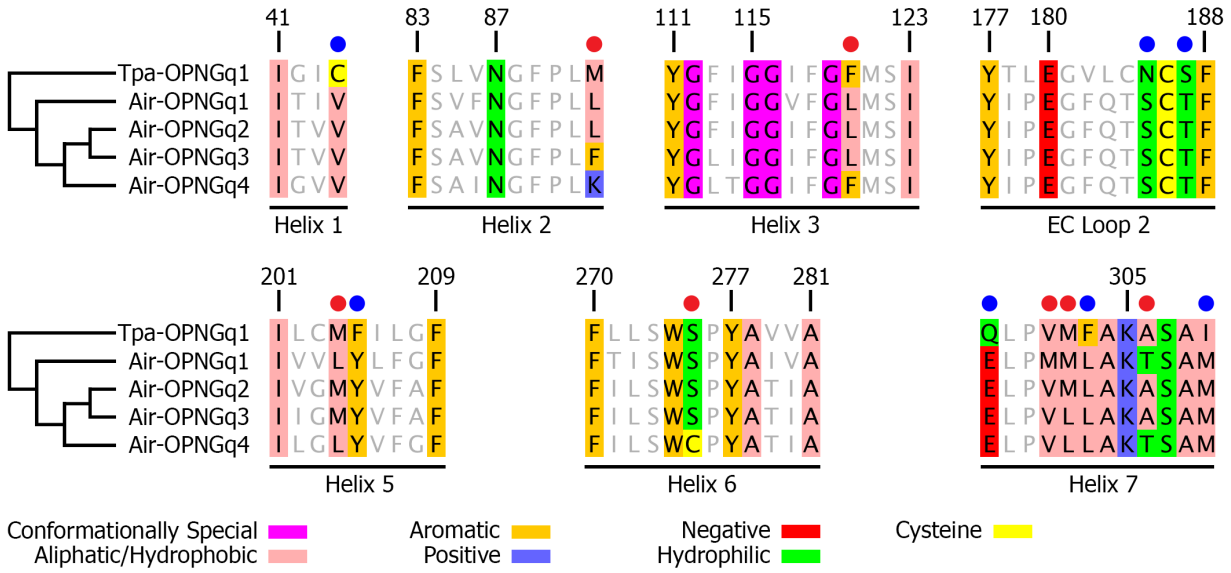


Figure 2-3 38 amino acid sites predicted to interact with chromophore in G_q-opsins. Predicted amino acids forming the chromophore pocket from a QM/MM model based on the Tpa-OPNGq1 crystal structure from Sekharan *et al*. [66]. We have inferred the putative chromophore pocket in scallop G_q-opsins by aligning all Air-OPNGqs against Tpa-OPNGq1. Blue dots indicated seven amino acid positions where all scallop G_q-opsins have the same residues and they differ from the Tpa-OPNGq1. Red dots identify the seven positions where amino acid residues differ among the four scallop G_q-opsins. Numbering is based on Tpa-OPNGq1. The residues are colored according to their physicochemical properties under the zappo color scheme in Jalview v2 [105]. Numbering of amino acid positions begins with the start codon (Met) of Tpa-OPNGq1; EC, extra cellular loop.

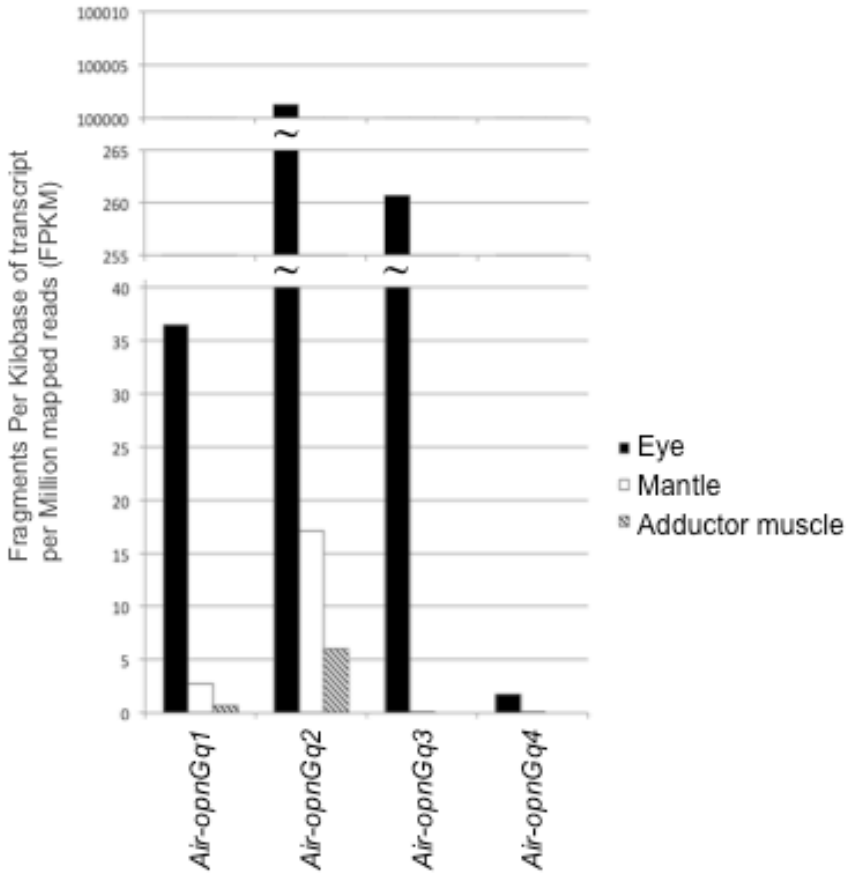


Figure 2-4 Expression profiles of scallop G_q-opsin genes across three tissues from a single light-treated animal. Gene-specific mRNA levels were quantified using RNA-seq of tissue-specific libraries: eye (black), mantle (white), and adductor muscle (striped). Expression levels are reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM).

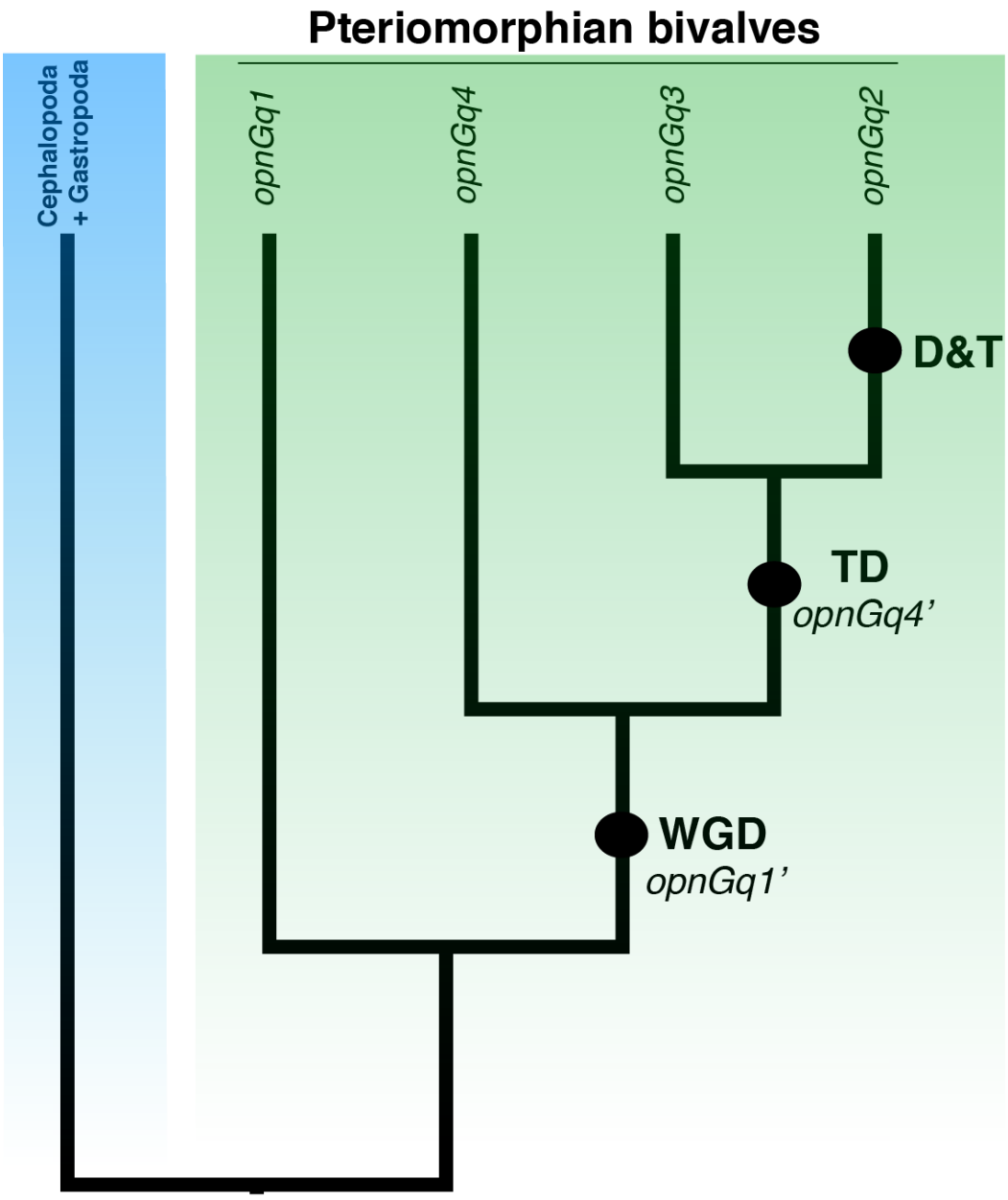


Figure 2-5 An evolutionary hypothesis describing G_q-opsin duplications in Pectinioidea. Circles represent duplication events. At least five bivalve families (Ostreidae, Pteriidae, Pectinidae, Spondylidae, Limidae) in three Pteriomorphia orders (Ostreida, Pterioida, Pectinoida [49]) possess a G_q-opsin gene homologous to *opn-Gq1*. One possible scenario is a whole genome duplication (WGD) event unique to the Pectinoida lineage [82] generated a second G_q-opsin copy (*opn-Gq1'* to *opn-Gq4*). *Opn-Gq4* under went a tandem duplication (TD), *opn-Gq4'*. A

final round of duplication generated *opn-Gq3* and *opn-Gq2*. The *opn-Gq2* copy was subsequently translocated (D&T) to another chromosome. While taxonomic sampling is not dense enough to determine whether these duplications are only in the family Pectinidae or in the order Pectinoida, current evidence supports the latter scenario as the Spondylidae and Limidae possess *opn-Gq2* homologs [49]. The other opsin paralogs (*opn-Gq3*, *opn-Gq4*) are either unsampled or may have been lost from these two families.

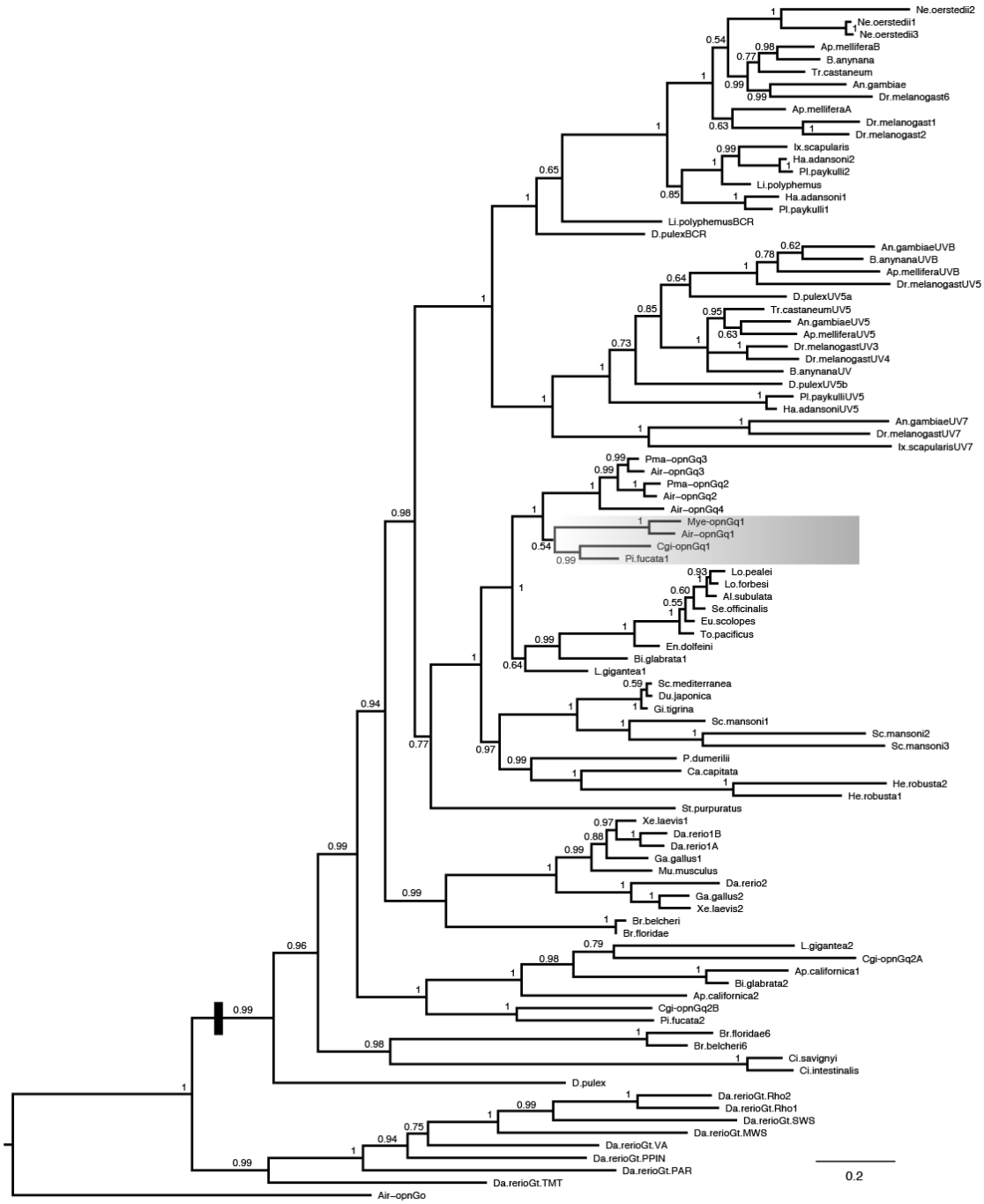


Figure 2-S1 Bayesian inference phylogram of G_q-opsins**.** The phylogenetic tree is based on 96 aligned amino acid sequences with scallop *Argopecten irradians* G_o-opsin as the outgroup. Support values at nodes are posterior probabilities >0.50. The grey box highlights a clade of

bivalve *opnGq1* not recovered in the ML analysis. A black bar indicates the monophyletic G_q-opsin clade.

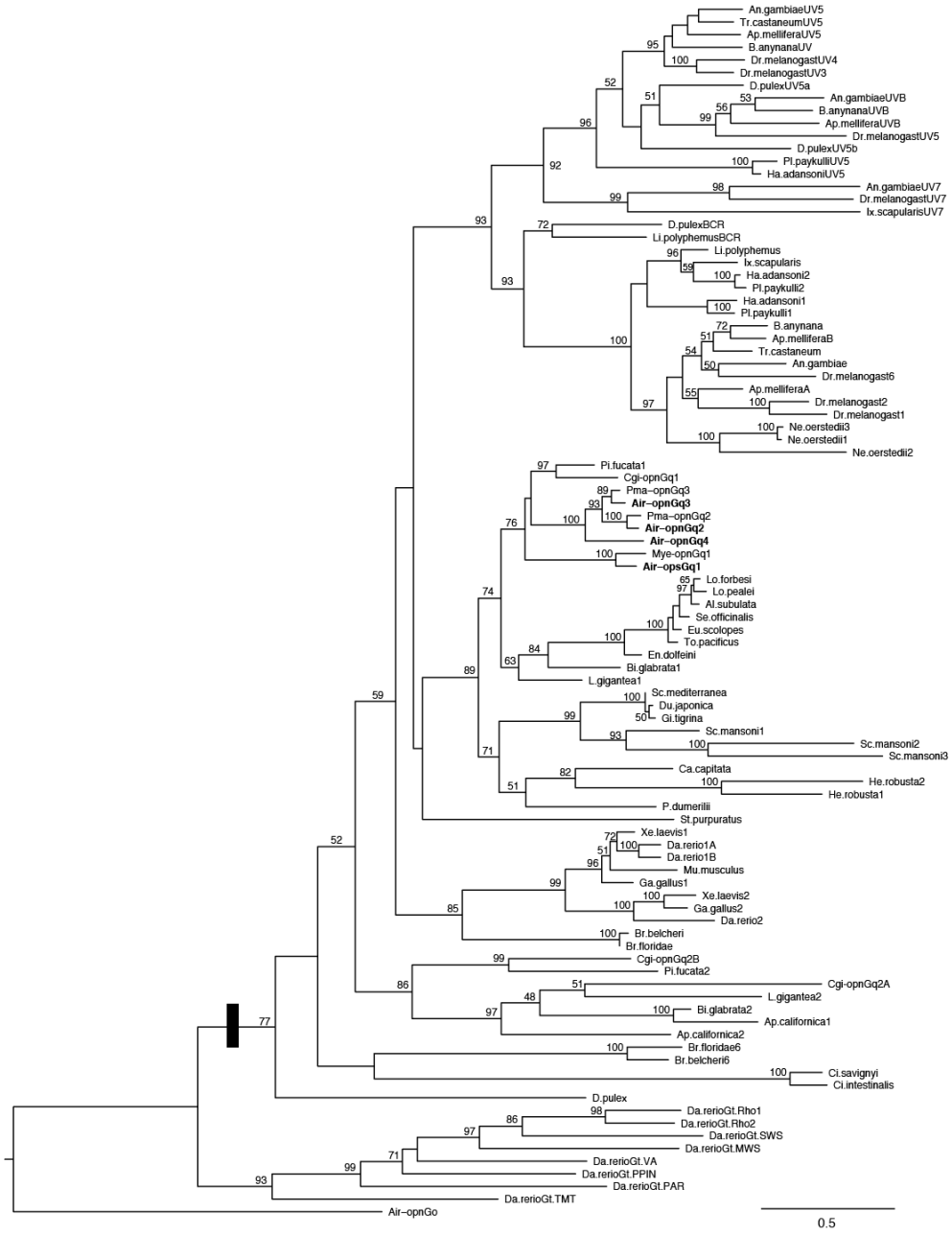


Figure 2-S2 Maximum likelihood phylogram of G_q-opsins. The phylogenetic tree is based on 96 aligned amino acid sequences with scallop, *Argopecten irradians*, G_o-opsin as the outgroup. Support values (>50%) of nodes were generated by 1000 bootstrap replicates in RAxML. A black bar indicates the G_q-opsin clade.

Figure 2-S3 Fifty base pair alignment of 5'- and 3'-UTRs from the four scallop G_q-opsins. Vertical lines represent the beginning and end of the coding region.

CHAPTER III

DIFFERENT EVOLUTIONARY TRAJECTORIES BETWEEN GQ-OPSIN PARALOGS SAMPLED ACROSS SCALLOPS

Anita Porath-Krause[1] and Jeanne M. Serb[1]

[1]*Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA*

Abstract

The retained products of gene duplication events are frequently studied to explore the sources of selection that are acting to retain that genetic material. Molecular evolution of invertebrate opsins and their rich history of lineage-specific duplication events have increasingly drawn the interest of evolutionary biologists because opsins have played an important role in expanding photoreceptive capabilities of organisms by altering what wavelengths of light are absorbed by photoreceptors (spectral tuning). Here, we examine a type of G-protein coupled opsin found in rhabdomeric cells, G_q-opsins, to determine how two paralogous clades differentiate after the duplication across a family of marine bivalves. We compare the rates of evolution between the paralogs and found that both clades are under purifying selection, and yet are maintained at rates that are significantly different. Only one amino acid position stands out as a strong candidate to link specific residue changes and explain the difference in evolutionary rates. We argue that the amino acid positions that are not commonly tested in studies of invertebrate spectral tuning may be critical to reveal important patterns of molecular evolution in G_q-opsins.

Keywords: opsin, evolutionary rate, spectral tuning, scallop

Introduction

Photoreception is important for the survival of many animals as it is a large component of many behaviors, including foraging (Dulai *et al.*, 1999; Gumbert, 2000), recognizing predators (Barcellos *et al.*, 2007) and conspecifics (Baracchi *et al.*, 2013), responding to shadows (Morton, 2008), and locating habitats (Hamilton & Koch, 1996). Gene duplication and subsequent functional divergence of genes involved in photoreception has played an important role in expanding photoreceptive capabilities of animals (Dulai *et al.*, 1999; Yuan *et al.*, 2007; Hunt *et al.*, 2009). As such, the genes essential for photoreception make an attractive model to explore molecular evolution.

Opsins are a type of G-protein coupled-receptor (GPCR) that when bound to a vitamin A-derived chromophore, form a light-sensitive photopigment. The ability of opsin to function as a photosensitive protein is dictated by specific amino acid motifs that allow the protein to bind to, and create a pocket around, the chromophore, such as the lysine-containing xAKxS motif in TM7 (reviewed in Rosenbaum *et al.*, 2009). Quantifiably measured as wavelength absorption, the wavelength at which the absorbance of light is the greatest (λ_{max}) is determined by either differences in chromophore type or the composition/shape of the chromophore binding pocket (Whitmore & Bowmaker, 1989; Asenjo *et al.*, 1994; Yokoyama *et al.*, 2008). Changes in the residues that form the chromophore binding pocket can alter the wavelength of light the photopigment is most sensitive (λ_{max} value), also termed as spectral tuning (Neitz *et al.*, 1991).

The opsin gene family went through many rounds of duplication early in metazoans to produce four opsin groups (Porter *et al.*, 2012), R-type, C-type, Group 4,

and cnidops (Plachetzki *et al.*, 2007), and subsequently multiple opsin subfamilies (Ramirez *et al.* 2016). Phylogenetic analyses demonstrate additional lineage-specific expansion of opsin subfamilies, which were caused by duplication events followed by neofunctionalization. For example, opsin duplication was the origin of color vision in fish (Yokoyama *et al.*, 2008; Rennison *et al.*, 2012), primates (Neitz *et al.*, 1991), and arthropods (Spaethe, 2004; Koyanagi *et al.*, 2008). In other systems, neofunctionalization of opsin followed by divergent selection at particular positions on the protein sequence resulted in spectral shifts, which are the increase or decrease in the λ_{max} value compared to the original value (Spady, 2005; Terai *et al.*, 2006; Frentiu *et al.*, 2007; Hoffmann *et al.*, 2007). Both experimental (Yokoyama *et al.*, 2008) and correlative studies (Spady, 2005; Hunt *et al.*, 2009) have identified a relationship between specific spectral tuning sites and adaptive evolution for specific light environments, so a higher rate of omega (the ratio of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site) is expected at sites involved in spectral tuning. Site-directed mutagenesis studies in vertebrates have revealed the importance of specific amino acid positions interacting with the chromophore and ultimately affect spectral tuning (Yokoyama & Radlwimmer, 1998; Takahashi & Ebrey, 2003), often called key residues (Murakami & Kouyama, 2008), key amino acids (Yokoyama, 2000), or key-sites (Carleton & Kocher, 2001). These positions are frequently used as the reference points when testing for changes in spectral tuning.

Gene duplication events supply the raw genetic material which can be exploited for new biological functions (Ohno, 1970). Many models propose that immediately after a duplication event, at least one gene copy will be under a period of relaxed selection

(reviewed in Taylor & Raes, 2004), which will then be followed by either positive selection (Ohno, 1970) or gene loss through pseudogenization (reviewed in Innan & Kondrashov, 2010). Relaxed selection can change the rate of evolution by reducing the strength of purifying selection, so that the number of non-synonymous substitutions per site exceeds the number of synonymous substitutions per site, maintaining higher genetic diversity than if under an increased strength of purifying selection. However, there is a lack of support for an all-encompassing rule regarding evolutionary rate for genes immediately post-duplication (Taylor & Raes, 2004). Differences in evolutionary rate cannot be explained exclusively by selection, but may be affected by several contributing factors including the mode of duplication (i.e. whole genome duplication, tandem duplication, etc.) (Yang & Gaut, 2011a), the length of the open reading frame (Marais & Duret, 2001), and the function or biological processes in which the duplicate is involved (Zuckerkandl, 1976).

Purportedly, the most influential contributing factors affecting evolutionary rate are the protein expression level and tissue-specificity (Yang & Gaut, 2011b). While these factors can affect evolutionary rate immediately after a duplication event, once the duplicates are fixed in a population, the rate of sequence evolution typically varies between the two duplicates due to different selective pressures such as behavior or environmental conditions. Frequently, duplicate genes exhibit low evolutionary rates due to purifying selection which is caused by strong functional constraints on the retained genes after the initial acceleration (during relaxed selection) (Jordan *et al.*, 2004).

Here, we examined the patterns of opsin sequence evolution using a new invertebrate model, the scallop (Pectinidae). Scallops contain two retinas, the proximal

and distal retina, each containing a different photoreceptor cell-type, rhabdomeric photoreceptor cells and ciliary photoreceptor cells, respectively. The proximal retina contains a G_q-protein coupled opsin gene (herein *opnGq* for the gene or the coding region, and OPNGq for the protein) and are frequently, but not exclusively, found to be involved in invertebrate photoreception, while the distal retina contains a G_o-protein coupled opsin gene which are more commonly associated with vertebrate visual capabilities. Evidence from morphological data of the scallop eye suggest that both retina are used for spatial vision (Speiser *et al.*, 2011) but further studies of comparative morphology indicate the proximal retina performs tasks that are used more in scallop species with the ability to swim (Speiser & Johnsen, 2008). A duplication of opsin genes was recently documented in the order Pectinoida which includes three bivalve families, Propeamussidae, Limidae, and Spondylidae (Serb *et al.*, 2013a). Phylogenetic analyses reveal an *opnGq* duplication event occurred in an oyster-Pectinoida ancestor, and the likelihood of retaining both genes is associated with either the presence of eyes and/or the degree of mobility (Serb *et al.*, 2013b). In this study, we tested whether the G_q-opsin paralogs in Pectinoida are under different types of selective regimes, which we define as the suite or aggregate of selective pressures, both organismal and environmental, on the population, using nucleotide sequence data we collected from 34 species. We hypothesized the G_q-opsin paralogs are under different selective regimes, which is attributed to stronger diversifying selection for specific nonsynonymous differences at those amino acid positions estimated to interact with the chromophore. We investigated which biological levels of organization that selection may be operating to explain the difference in selective regimes. To identify the biological level of selection, we tested for

species-specific episodes of selection, selection on individual sequences within each G_q-opsin paralog clade, and site-specific selection. We applied a protein folding model developed from crystallography data of the squid OPSGq ("squid rhodopsin"; Murakami and Kouyama 2008; Shimamura *et al*. 2008) to target amino acid positions predicted to be important for spectral tuning through their interaction with the chromophore. To identify the source of the selection, we tested for a correlation between G_q-opsin sequence and two characters, behavior class or photic environment that may have an effect on opsin sequence evolution. We predicted that at least one prospective spectral tuning position (Sekharan *et al.*, 2012), or multiple positions within the transmembrane regions, to be under positive selection and responsible for the rate differences between *opnGq1* and *opnGq2*.

Methods

a) Study system

Scallops are multi-eyed, marine bivalves belonging to the family Pectinidae. Each species can be placed into one of six life habit classes that are defined mainly by how the adult animal attaches to a substrate and its ability to move (see Table 1 in Alejandrino *et al.*, 2011). Ranging from most mobile to least mobile, these classes include: gliding, free-living, byssal attaching, recessing, nestling, and cementing (see Alejandrino *et al*., 2011 for details). Gliding and free-living species are able to swim but do not attach to their substrate. Recessing species do not attach, either, but in place of swimming, these species will partially or completely bury themselves in sand or sediment. Nestling and cementing species will secrete a byssus to attach to objects and become permanently

attached to the object. Byssal attaching species will secrete a byssus to attach to an object but can detach and reattach if necessary. Scallops are found in nearly every ocean, dwelling in a wide range of depths, from 1 to 7000 meters (Dijkstra, 1991; Waller, 1991). The photic environment can vary depending on water clarity, depth, and light intensity (Lythgoe, 1968; Jerlov, 1976), so each scallop species may be exposed to different regions of the visual light spectrum.

b) G_q-opsin isolation and sequencing

In this study, we examined differences in the composition and the rate of nucleotide substitution of the coding region of two G_q-opsin paralogs, *opnGq1* and *opnGq2* for two reasons. First, both these paralogs have the highest level of gene expression in the eyes and the photosensitive mantle tissue (Porath-Krause *et al.*, 2016). Second, *opnGq1* and *opnGq2* were isolated from eight scallops species and three closely allied bivalve families, Propeamussidae, Limidae, and Spondylidae (Serb *et al.*, 2013b), making it likely that both opsins would be present in other species across the Pectinoida. We sampled 31 scallop species across the Pectinidae to represent different life habits (Figure 1 in Alejandrino *et al.*, 2011). A single species from each of the three closely allied bivalve families were also selected. While some scallop species are available through aquaculture, many species are difficult to collect and obtain quality RNA samples. Additionally, scallop *opnGq2* does not contain introns and only one intron is found in *opnGq1,* so we maintained a wider breadth of sampling across the phylogeny by isolating genomic DNA from preserved adductor muscle tissue. Ethanol-preserved

tissues from the Pectinoida were obtained from museum collections or provided by colleagues (see Table 1 in Alejandrino *et al.*, 2011).

Total genomic DNA was extracted from ethanol-preserved samples following the methods described by Puslednik and Serb (Puslednik & Serb, 2008). Polymerase chain reaction (PCR) was performed using published gene-specific primers and methods described in Serb *et al*. (2013) (Table 3-1). We amplified a 489 base pair region that includes transmembrane (TM) 4 through TM7. This region represents approximately 35% of the entire coding region and contains the lysine motif that binds the chromophore. We chose this region because it captures 26 of the 38 positions considered putative spectral tuning sites (Sekharan *et al.*, 2012); and specifically, this region includes the only three positions identified to have nonsynonymous differences between G_q-opsin paralogs in *A. irradians* (Porath-Krause *et al*., 2016). PCR products were cloned into a TOPO TA vector (pCR2.1-TOPO) that was used to transfect *E. coli* following the manufacturer's instructions (TOPO TA Cloning Kit). Colonies were blue-white screened and at least ten colonies containing recombinant DNA were selected. We amplified the recombinant DNA from each colony using M13 forward and M13 reverse primers according to manufacturer's instructions (TOPO TA Cloning Kit, Invitrogen). We confirmed that the plasmid contained an insert of the target size using a 1% agarose gel electrophoresis. Clones with the correct insert size were Sanger sequenced with an ABI 3730 Capillary Electrophoresis Genetic Analyzer at the Iowa State University DNA Sequencing Facility. The resulting DNA sequences were translated and manually checked for stop codons and the presence of the xAKxS motif in TM7 indicating chromophore-binding capability of the translated protein. DNA sequences that did not meet these requirements were

considered nonfunctional pseudogenes and were removed from further analyses. Sequences were blasted in the NCBI GenBank database using blastn to confirm gene identity as either *opnGq1* or *opnGq2* (described as opnGqB and opnGqA, respectively, Serb *et al.*, 2013b). All newly generated sequences were deposited in NCBI GenBank database (accession numbers to be entered after submission) (Table 3-2).

c) Predominant sequence variants

Because the shape of the chromophore-binding pocket can affect the absorbance peak (λ_{max}) of the photopigment, we were interested in non-synonymous amino acid changes that may influence tertiary structure of the opsin protein. For each paralog, we retained all opsin sequences that represent intraspecific variation. These nucleotide sequences were translated into amino acids for a total of 90 OPNGq1 and 101 OPNGq2 amino acid sequences among the 34 species. For each species, we then determined a predominant sequence variant for OPNGq1 and OPNGq2. We define the predominant sequence variant as either 1) the only amino acid sequence for that species or 2) the most common amino acid sequence for that species. All protein sequences are numbered according to Figure 2-1 (Porath-Krause *et al*., 2016), where the first amino acid aligns with the start codon of the squid- G_q-opsin.

A maximum likelihood topology of G_q-opsin predominant amino acid sequence variants was estimated based on aligned amino acid sequences. Thirty-four taxa from the Superorder Pectinoida are represented by two paralogs, OPNGq1 and OPNGq2. Sixty-seven predominant sequence variants (34 OPNGq1 + 33 OPNGq2) with *Todarodes pacificus* (squid) rhodopsin as a non-bivalve mollusc outgroup (NCBI accession

#X70498) using JTT+Gamma (Jones *et al.*, 1992) (the best model of selection determined in ProtTest v. 2.4) (Abascal *et al.*, 2005), in PhyML version 3.0 (Guindon & Gascuel, 2003) through Geneious version Pro 5.6.7 (Drummond *et al.*, 2011). Support values (>50%) of nodes were generated by 1000 bootstrap replicates.

d) Tests of selection across paralogs

Comparing the molecular evolutionary trajectories of two duplicates can reveal the patterns of selection on the genes. An evolutionary trajectory is an accumulation of mutations over time that will produce a phenotype. If the mutation rate of a nucleotide sequence is compared to the ancestral sequence or duplicate gene mutation, the type of selection influencing an evolutionary trajectory can be estimated. Selective regimes, such as positive, purifying, and neutral selection can be estimated using the ratio of nonsynonymous to synonomous changes between nucleotide sequences because the relative rate of mutation in nucleotide sequences may produce observable patterns (Ohta, 1994; Lemey *et al.*, 2007; Light & Hafner, 2008; Jenkins *et al.*, 2014). Quantifying omega, the ratio of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site (ω = dN/dS rate ratio) is a common way to test the extent of selection on sequence evolution.

We use several approaches to estimate the ratio of substitution rates at non-synonymous and synonymous sites and compare selection pressures acting within and between *opnGq1* and *opnGq2*. For the first approach, we determined whether the rate of evolution as measured by omega (ω = dN/dS) was faster in one G_q-opsin paralog compared to the other using the branch model implemented in the CODEML program of

the software package PAML version 4.7 (Yang, 2007) (http://abacus.gene.ucl.ac.uk/software/paml.html). The branch model detects selection acting on particular lineages of a tree by allowing the rate of evolution to vary among branches (Yang, 1998). Selective constraints following duplication events can be estimated by comparing the likelihood values of two trees with lineages containing different omega values.

Three types of data are needed to estimate branch or lineage-specific omega ratios using CODEML. These data include a nucleotide alignment, a topology based on the sequence alignment (a gene tree), and CODEML code specifying the parameters to test specific lineages for selection. We aligned the nucleotide sequences for 67 predominant sequence variants (34 OPNGq1 + 33 OPNGq2) with *Todarodes pacificus* (squid) rhodopsin as the outgroup (NCBI accession #X70498) using the default settings in MAFFT v7.017 (Katoh *et al.*, 2002) in Geneious version Pro 5.6.7 (Drummond *et al.*, 2011). From the alignment, a gene tree was constructed using HKY85 model (determined in jModelTest 2.1.1 (Posada, 2008) as the best-fit nucleotide model of evolution) (Hasegawa *et al.*, 1985) with PhyML version 3.0 (Guindon & Gascuel, 2003) in Geneious version Pro 5.6.7 (Drummond *et al.*, 2011) using 1000 bootstrap pseudoreplicates (Figure 3-S1). Based on previous phylogenetic results in Serb *et al.* (2013b), we expected the *opnGq1* and *opnGq2* gene sequences we collected would form two monophyletic clades.

Once a nucleotide alignment and corresponding gene tree were built, we estimated the rate of ω in each gene lineage, *opnGq1* and *opnGq2*. We tested for lineage-specific episodes of selection by designating lineages as either foreground or background

branches. Foreground branches have a positive selection model applied (omega = X), while all other branches are considered background branches and have a null model applied. We performed seven tests (T0 through T6) to estimate the rate of omega on the foreground and background branches including: a null model (T0) which kept the selection pressure homogenous over the tree; Model T1 which allowed an episodic change in selection pressure in *opnGq2* but only in the branch that immediately follows the gene duplication event; Model T2 where there was a long term shift in selection pressure in *opnGq2* only, but *opnGq1* remains subject to the ancestral level of selection pressure; Model T3 where there is a long term shift in selection occurs in both *opnGq2* and *opnGq1*, where those lineages differ from each other and from the ancestor; and three alternative models which are the reverse complement of T1, T2, and T3 using *opnGq1* as the foreground branch (T4, T5, and T6) (Figure 3-S1). In CODEML, the model number was specified as M0 or M2, which represent different omega ratio for the assigned branches. If a M of 0 is assigned, there is one omega ratio assigned for all branches. If a M of 2 is assigned, then more than one omega ratio can be assigned to the branches. Tests T1 though T6 had M2 model, while test T0 had M0 model (Table 3-3). All tests had a codon frequency value of F3x4 (CodonFreq=2), and an initial omega value of 0.2, which were estimated empirically from the nucleotide sequence data. Each of the seven tests produced an omega value for each lineage and a *ln* value, which is a likelihood estimate of that tree under the specified selection pressures. The likelihood values from the seven tests (T0-T6) were then used to address the hypotheses in Table 3-4. Hypothesis A: The mutation rate of *opnGq2* has changed relative to *opnGq1*. Hypothesis B: A burst of positive selection for functional divergence occurred following the

duplication event that gave rise to *opnGq2*. Hypothesis C: There was a long-term shift in selective constraints following the duplication event that gave rise to *opnGq2*. Hypothesis D: The mutation rate of *opnGq1* has increased relative to *opnGq2*. Hypothesis E: A burst of positive selection for functional divergence occurred following the duplication event that gave rise to *opnGq1*. Hypothesis F: A long-term shift in selective constraints followed the duplication event that gave rise to *opnGq1*. The log likelihood values for each test were then compared using a likelihood ratio test in which the significance is determined from the χ^2 distribution with one degree of freedom.

To demonstrate if selection has influenced a shift away from a normal null distribution, I compared the observed patterns to empirically generated random distributions. The nucleotide sequences from the 67 bivalve species were randomized while keeping the species' names fixed on the tree to create a null distribution of omega values for *opnGq1* and *opnGq2*. Each new randomly chosen set of sequences was then used to perform an identical test to Model T2. The resulting omega values were then used to create a normal distribution with a mean and standard deviation for each Gq-opsin paralog. The omega values obtained from the original, non-randomized sequences were then compared to the distribution.

d.1) Testing for selection among sequences

We tested for selection acting on individual sequences within each G_q-opsin clade by comparing the relative abundance of nucleotide substitutions between two sequences. Using the nucleotide alignment described in the *Tests of Selection* section, we implemented the codon based Z-test (Nei & Gojobori, 1986) available in MEGA version

6.06 (Tamura *et al.*, 2013). The codon based Z-test compares the relative abundance of synonymous (dS) and nonsynonymous (dN) substitutions that occur across the sequences in a pair-wise format (Nei & Kumar, 2000). dN and dS values were estimated at each site using the Nei-Gojobori method (Nei & Gojobori, 1986), and then variance was calculated across the sequence. A Z-score was obtained to test the null hypothesis of neutrality, so that H0: dN = dS. If the null hypothesis was rejected at significance level of 0.05, then we can infer either diversifying selection, dN > dS, or purifying selection, dN < dS, was acting on that sequence.

d.2) Testing for site-specific selection

Because tests for selection across a sequence can mask selection happening at specific amino acid sites, we tested for selection on individual codons. We tested for purifying or diversifying selection at each codon using the program HyPhy Version 2.2.1 (Pond & Frost, 2005a) on the DataMonkey Webserver (Delport *et al.*, 2010). Because the two paralogs have only approximately 50% amino acid identity, we examined the nucleotide sequences for each G_q-opsin paralog separately to account for the age of the copies which likely have had multiple substitutions at a site. We trimmed a comprehensive scallop species tree (Alejandrino *et al.*, 2011) to include only those taxa representative of the *opnGq1* and *opnGq2* nucleotide alignments. Then, each trimmed-tree topology was used with the corresponding nucleotide alignments to complete the HyPhy analyses. Selection was estimated by determining the rate of substitution at each site, then using a likelihood ratio test to see if the rates were significantly different between nonsynonomous and synonomous sites. A fixed effects likelihood model (FEL) (Pond & Frost, 2005b) was

chosen over the random effects likelihood model (REL) (Pond & Frost, 2005b) to detect amino acid sites under selection because FEL has been shown to have a lower rate of false positives than REL when analyzing small data sets (Pond & Frost, 2005b).

To determine if the amino acid sites found to be under positive selection using the branch-site model in CODEML (PAML) (Yang, 2007) (http://abacus.gene.ucl.ac.uk/software/paml.html) differ from those found using the fixed effects likelihood model (FEL) in HyPhy (DataMonkey) (Pond & Frost, 2005; Delport *et al.*, 2010), I used branch-site model A in CODEML to estimate ω for each site between a foreground and a background branch. A log-likelihood ratio test was then used to determine if the model is a better fit than the null model where there is uniform selection pressure among sites (ω can vary among sites, but ω = 1 for the background and foreground branches). The Bayes Emperical Bayes method was then used to calculate the posterior probability at each site being positively selected. Two analyses were performed, one with *opnGq1* as the assigned foreground and one with *opnGq2* as the assigned foreground.

e) Identifying amino acid positions that affect evolutionary rate

It has been shown experimentally that specific amino acid positions dictate the shape of the chromophore pocket through differences in the biochemical properties of the amino acid residues; specifically, a change of charge (polar vs non-polar) and a gain or loss of a hydroxyl group can shift the λ_{max} value at those specific amino acid positions (Asenjo *et al.*, 1994; Yokoyama *et al.*, 2008). While evolutionary rates can be informative regarding selection, we wanted to explore if the difference in evolutionary

rate reflects differences in the amino acid biochemical properties between OPNGq1 and OPNGq2. To explore the source of evolutionary rate variation among genes, we used the G_q-opsin maximum likelihood topology (Figure 3-S1), the nucleotide alignment described in the *tests of selection* section, above, and the program DIVERGE v. 3.0 (Gu & Vander Velden, 2002). DIVERGE was used to estimate Type I and Type II source variation at each codon across the *opnGq1* and *opnGq2* nucleotide alignment. Type I source variation is defined as a highly conserved amino acid pattern within one clade and a highly variable amino acid pattern in the other clade. Type II source variation is defined as a very conserved amino acid pattern in each clade, but the biochemical properties are very different between the two clades (Gu, 2001). Biochemical properties are defined as hydrophilic, hydrophobic, positively charged, and negatively charged amino acid groups. Evolutionary rate is detected by comparing rate correlation between two clades using either a two-state model of equal rates between clades or unequal rates between clades. To test for differences in evolutionary rates between G_q-opsin paralogs, we identified *opnGq1* and *opnGq2* as two separate clades and compared rate correlation between two clades (*opnGq1* clade versus *opnGq2* clade). A maximum likelihood model was used to estimate the probability of rates being statistically independent between each clade and is estimated at each codon. If a rate is independent, then the source of variation is categorized as either Type I or Type II.

We used ancestral state reconstruction to explore patterns of independently evolved, site-specific selection to explain the difference in selective regimes. For each G_q-opsin paralog, we mapped the nonsynonymous positions onto the trimmed Pectinidae phylogeny and then estimated the amino acid ancestral states using maximum likelihood

in MEGA (Tamura *et al.*, 2013) based on the best model of selection, JTT+Gamma (Jones *et al.*, 1992) determined in ProtTest v. 2.4 (Abascal *et al.*, 2005). Any nonsynonymous positions containing amino acids that differed from the ancestral amino acid were of interest because those changes could indicate that a particular amino acid experienced positive selection.

f) Testing for an association between behavior or environmental light conditions and G_q-opsin sequence variant

We tested if specific sequence variants were correlated with a scallop's behavior or photic environment using the method by Pagel and Meade (2006) of discrete characters in BayesTraits V2 (Pagel, 1994). We compared the fit of dependent and independent models using the Discrete function with binary traits (presence/absence) for each of the following characters: behavior, depth of habitat, sequence variant for OPNGq1, and sequence variant for OPNGq2. The independent model assumes two traits evolve independently while the dependent model assumes the rate of change in one character is dependent on another. A maximum likelihood approach was implemented with 1000 bootstrap support to calculate the log-likelihoods for each model, and then compared dependent and independent models using a likelihood ratio test and Akaike Information Criterion (Akaike, 1974). The correlation was considered significant if the likelihood ratio was less than 0.05 for $P(X^2_{df=4})$. A Bayesian approach was also implemented using Markov Chain Monte Carlo in BayesTraits V2 (Pagel & Meade, 2006) to compare the dependent and independent models by calculating the harmonic mean of the likelihoods. The calculation was repeated several times and estimated using

100,000,000 runs. The log Bayes Factor (BF) was calculated by taking twice the difference of the log of the independent and dependent models. We only accepted the strictest category of very strong evidence (log BF >10) as support for a correlation between the trait and sequence variant.

For each analysis, each of the six representative sequence variants (five predominant sequence variants clades (A-E) + one representative that encompasses all remaining unique sequences) (refer to Figure 3-1) was tested separately to determine if any one representative sequence variant was correlated with a specific trait. For example, amino acid sequences were assigned a "1" if it belonged to sequence variant A, which was found in 20 species, and a "0" for the remaining unique sequence variants. In the next analysis, amino acid sequences were assigned a "1" if it belonged to sequence variant B, and "0" for the remaining sequence variants. Each of the six behaviors and four depth categories were analyzed individually, as well, using the same binary presence/absence approach. Below the water's surface, the spectral range of light becomes restricted such that the available wavelengths of light decrease as the depth increases (i.e. blue wavelengths of light can penetrate deeper waters while red wavelengths of light can only penetrate shallow waters); so, depth categories are defined based on the table from Jerlov (1976; Figure 132, page 198) assuming the best conditions with the clearest of ocean water (Type I; Jerlov, 1976), our depth categories include shallow ($\leq$ 20m), medium ($\geq$ 21m and $\leq$ 65m), deep ($\geq$ 66m and $\leq$ 110m), and very deep waters ($\geq$ 111m). A range depth (in meters) was identified for each species from multiple literature sources (Table 3-S1). Because scallop species can have a broad geographic distribution, we restricted the depth range to include only populations of species collected

from the same biogeographic province as those sampled for molecular data. Additionally, we discarded depth values for species collected by trawling, a fishing technique that drags a net behind a boat, because during dragging multiple depths are sampled.

Results

We were able to isolate both *opnGq1* and *opnGq2* from 30 scallop and three outgroup species. The exception is *Crassadoma gigantea,* which does not appear to contain a functional *opnGq2* sequence because all clones that were sequences contained multiple stop codons and did not have the lysine-containing xAKxS motif in TM7. There were a total of 34 predominant sequence variants for OPNGq1 and 33 predominant sequence variants for OPNGq2. The maximum likelihood topology built with G_q-opsin predominant sequence variants is not congruent with the scallop species tree (Alejandrino *et al.*, 2011). For example, the three non-Pectinidae families used as outgroup taxa in the species tree are nested within the G_q-opsin clades. This suggests the selective pressures shaping the evolution of G_q-opsin paralogs are more similar than the phylogenetic relationship across these bivalve species. Maximum likelihood analyses of all G_q-opsin amino acid sequence variants resulted in an amino acid tree, which showed OPNGq1 and OPNGq2 form two monophyletic clades. The relationships within each G_q-opsin clade vary greatly from one another, especially with the number of identical predominant sequence variants. In the OPNGq1 clade, the majority of species had 20 predominant sequence variants form one large group of identical sequences, while within the OPNGq2 clade there are four groups of identical predominant sequence variants. Within the four

groups of identical sequence variants found within the OPNGq2 clade, two groups contain two identical predominant sequence variants each, and two groups contain three identical predominant sequence variants each (Figure 3-1). There are a total of 14 unique (34 total – 20 identical) predominant sequence variants representing OPNGq1 clade and there are 23 unique (33 total – 10 identical) predominant sequence variants representing OPNGq2 clade.

Variable rates of evolution between opnGq1 and opnGq2

Under maximum likelihood, *opnGq1* and *opnGq2* nucleotide sequences form two well-supported monophyletic clades (bootstrap support = 83%) (Figure 3-S1). Using this phylogeny, we examined the difference in evolutionary rates between *opnGq1* and *opnGq2* nucleotide sequences. Across the G_q-opsin sequence topology, the null model, T0, had an average ω of 0.071 (Table 3-3 and Figure 3-S1). Each alternative model (T1, T2, T4, and T5) was a significantly better fit to the data than the null model (T0) (Table 3-4), supporting a non-uniform rate of evolution occurred after the duplication event that gave rise to *opnGq1* and *opnGq2*.

Duplicate genes frequently exhibit low evolutionary rates due to purifying selection, which is caused by strong functional constraints on the retained genes. We measure the strength of selection on each paralog and estimate if the selection pressure is continuous (a long-term shift) or a point(s) in time (episodic). When estimating if a change in selection pressure was more likely episodic (T1 or T4) or the result of a long-term shift (T2, T3, T5, or T6), we found a long-term shift in selection pressure was more likely (Table 3-4). Hypothesis B (T0 vs T2) and E (T0 vs T5) which compare a

homogenous selection pressure across the G_q-opsin sequence topology to a long term shift in selection pressure in one G_q-opsin paralog, had the highest difference in likelihood values (14.892 and 18.064, respectively) and the most significant values. In both comparisons, the rate of change was nearly twice that in *opnGq1* ($\omega = 0.119$) compared to *opnGq2* ($\omega = 0.056$). Even when we compared the T2 to T3 (Hypothesis C: $2\Delta l = 7.994$; $P(X^2_{df=1}) = 0.0047$) and T5 to T6 (Hypothesis F: $2\Delta l = 4.822$; $P(X^2_{df=1}) = 0.0281$) to test if there was a difference in selective pressure for both *opnGq1* and *opnGq2* following gene duplication, omega was nearly two times larger in *opnGq1* than *opnGq2* (Table 3-4). This suggests that *opnGq1* is evolving faster than *opnGq2*.

Omega values obtained from the *opnGq1* randomized sequences was $0.0221 \leq \omega \leq 0.0588$ with a mean $\mu = 0.0411$ and a standard deviation $\sigma = 0.0105$. *opnGq2* randomized sequences had a smaller range of omega values ($0.0307 \leq \omega \leq 0.0524$) but a very similar mean $\mu = 0.0415$. The standard deviation for *opnGq2* was smaller than *opnGq1* with $\sigma = 0.0063$. The omega value ($\omega = 0.056$) obtained from non-randomized *opnGq2* sequences under the long-term shift test in selection pressure is within three standard deviations, or 99.7%, of the mean. The omega value ($\omega = 0.119$) for *opnGq1*, however, falls outside three standard deviations of the mean ($\mu + 3\sigma = 0.0726$). This suggests *opnGq1* is under weaker purifying selection than expected from a normal null distribution of Gq-opsin nucleotide sequences sampled here.

Most sequences are under purifying selection

We explored the role of selection between sequences by comparing the number of synonymous and nonsynonymous substitutions using a pairwise comparison. Using

the codon-based Z test, we found no evidence that any sequence was under diversifying selection ($p > 0.05$) (Table 3-S2); however, the probability of rejecting the null hypothesis of neutral selection was in favor of the alternative hypothesis ($dS > dN$) (Table 3-S3). Greater than 80% of *opnGq2* pairwise sequence comparisons were found to be under purifying selection, which is not surprising because the rate of evolution within the *opnGq2* clade was quite low. Interestingly, over 50% of pairwise sequence comparisons from *opnGq1* were not found to be under significant purifying selection; however, most of those comparisons belong to nucleotide sequences that represent identical predominant sequence variants. We would expect identical or nearly identical nucleotide sequences to have equal dN/dS ratios and thus, not be under significant purifying selection. Those sequences that are not identical at the amino acid level and do not belong to the clade comprising sequence variant A demonstrate a significant value for purifying selection.

opnGq2 *has a higher number of codon positions under purifying selection than* opnGq1

Tests for site-specific selection reveal no clear patterns of purifying or diversifying selection. In the HyPhy results of the *opnGq1* alignment, 28% of codon positions were identified to be under purifying selection while one codon position, position 246, was identified to be under diversifying selection. This position is located in transmembrane VI, but is not homologous to any of the 38 putative spectral tuning sites in squid- G_q-opsin (Sekharan *et al.*, 2012). In contrast, 67.3% of codon positions were identified to be under purifying selection in *opnGq2,* while no positions were identified to

be under diversifying selection. Six of the *opnGq2* and four of *opnGq1* positions under purifying selection represent positions identified as putative spectral tuning sites in squid-G_q-opsin (Sekharan *et al.*, 2012) (Figure 3-2).

Using the branch-site method in CODEML, neither of the models tested was significantly better than the null model of neutral evolution ($\omega = 1$) (Table 3-S4). According to Bayes Emperical Bayes, however, there were 13 positions with a posterior probability p >50% for positive selection for *opnGq1*. Seven of which are considered significant with p >95% (* in Table 3-S4). These seven sites are located within transmembranes IV and V. Position 246 is the same site identified to be under positive selection using HyPhy in DataMonkey, but this was the only site found to be under positive selection in either *opnGq1* or *opnGq2* using HyPhy. In *opnGq2*, three positions were identified in CODEML to have a posterior probability p >50% for positive selection but none of which are considered significant (p >95%).

To detect the site-specific source of variation (Type I vs. Type II source variation) along the gene tree built using the *opnGq1* and *opnGq2* nucleotide sequences, we explored the source of evolutionary rate variation among clades using DIVERGE. Type I source variation is a highly conserved amino acid pattern within one clade and a highly variable amino acid pattern in the other clade, while Type II source variation is a conserved amino acid pattern in each clade, but the biochemical properties are very different between the two clades (Gu, 2001). The results suggested no specific position is solely responsible for the difference in evolutionary rate. From the nucleotide alignment, 34.83% of sites have no differences within or between *opnGq1* and *opnGq2*. When translated, a total of 70 sites were different between the OPNGq1 and OPNGq2 with a

higher proportion of conserved differences (0.5571) than differences in the biochemical properties (0.44285). For example, we saw a conserved difference at position 204 with a non-polar Leu across sequences in OPNGq1 and a non-polar Met across sequences in OPNGq2. Position 308 represents a difference in polarity between the OPNGq1 (polar Ser) and OPNGq2 (non-polar Ala).

Based on site-specific predictions, Type I posterior probability values were less than 0.50 at all positions, indicating no specific positions are solely responsible for the difference in evolution rate. Type II functional divergence identified several (68 of the 156) positions that are contributing the most to those radical changes in amino acid property groups between the G_q-opsin paralogs. Approximately 87% of these positions are within the transmembrane regions of the protein. This pattern could indicate a difference in λ_{max} values as those amino acid positions causing wavelength absorption differences are most often found within transmembrane regions (Yokoyama *et al.*, 2008). Studies have shown nonsynonymous differences frequently occur outside the transmembrane region upstream and downstream of the transmembrane regions, as well, but rarely affect the λ_{max} values because these positions do not interact with the chromophore (Shi & Yokoyama, 2003; Yokoyama *et al.*, 2008; but see Smith & Carleton, 2010). Notably, of the 70 positions that were identified to be under the Type II functional divergence, only two positions, 306 and 308, are homologous with spectral tuning sites modeled in the squid rhodopsin (Sekharan *et al.*, 2012). These two positions are located within TMVII, one and three amino acids downstream of the lysine-binding site.

Positions that are not considered putative spectral tuning sites explain the difference in evolutionary rates

The two G_q-opsin gene duplicates are very distinctive in their composition of sequence variants. OPNGq1 has one identical predominant sequence variant, which is possessed by 20 scallop species as seen in clade A. In contrast, ten species comprise clades B, C, D, and E, each of which contains identical OPNGq2 sequence variants. Focusing on putative spectral tuning sites, we found that a lack of nonsynonymous differences fail to explain the significant difference in evolutionary rate between the two gene duplicates (OPNGq1 vs OPNGq2), and also fails to explain why multiple unrelated species possess an identical sequence variant. Table 3-5 represents amino acid sites modeled to form the chromophore pocket (Sekharan *et al.*, 2012) and vary in amino acid composition. The remaining 17 amino acid sites estimated to form the chromophore pocket were not included in the table because there were no differences within or between OPNGq1 and OPNGq2. Amino acids are relatively conserved at the ten sites shown in the table. Only two positions, S306A and S308A, are under Type II divergence, and show a radical difference in the amino acid group between the majority of OPNGq1 and OPNGq2 predominant sequence variants. All other positions have an identical amino acid between OPNGq1 and OPNGq2. There are a few instances of a radical amino acid group difference within a duplicate, but these differences, Y177H, S185P, F209Y, L303S, and A306S, are found between OPNGq2 predominant sequence variants (B-E clades). These few differences do not explain the significant difference in evolutionary rate between the two paralogs.

When we expanded our focus to include positions that are not considered putative spectral tuning sites, we found multiple amino acid positions that could explain why there is a difference in evolutionary rates. Concentrating on differences in amino acid positions between sequence variants that comprise the B, C, D, and E clades, we identified 13 positions with nonsynonymous differences between these four clades (ten sequences). In Table 3-6, we highlight seven of these positions, 195, 214, 221, 247, 265, 288, and 297, which represent nonsynonymous differences across all OPNGq2 sequence variants that include at least one radical change in amino acid property groups. All seven of these positions are under Type II Divergence and three positions, A195T, R247E, and P288D, show a radical difference in the amino acid group between the majority of OPNGq1 and OPNGq2 predominant sequence variants. We also included one amino acid position from OPNGq1 that represents an amino acid position with less than 75% of amino acids sharing the same residue. All other OPNGq1 amino acid positions are synonymous at greater than 75% of the residues at that position. At position 217, multiple species contain nonsynonymous differences that are considered radically different from the majority. This position is under Type II Divergence and while OPNGq2 is under purifying selection, OPNGq1 is not under purifying selection at this position. Ancestral state reconstruction of the OPNGq1 and OPNGq2 amino acid sequences revealed putative spectral tuning sites had no radical amino acids at the most recent common ancestor compared to the amino acids of the extant species. When ancestral amino acids were compared to those positions identified to include radical amino acid differences (Table 3-6), multiple positions appeared to contain an independently evolved amino acid across one or two species at that position (Figure 3-2).

Two positions, 196 and 217, have multiple radical amino acid difference across species at the most recent common ancestor, but there is a clear pattern across the ancestral amino acids that reflects the pattern of the Pectinidae species tree. One position, 288, is of particular interest because all six gliding species, which independently evolved four times (Alejandrino *et al.*, 2011), maintain a radical difference at that position D288N compared to all other species sampled here (Table 3-6). The ancestral state reconstruction reveals the most recent common ancestor maintained an acidic polar aspartic acid (D) at this position, with the exception of the most recent common ancestor of *Ylistrum balloti* and *Ylistrum japonicum*, which maintain a neutral, polar asparagine (N). However, the most recent common ancestor of these *Ylistrum* sister species and *Decatopecten plica* is aspartic acid. These patterns suggest the transition from aspartic acid (D) to asparagine (N) occurred independently three times across scallops and only in lineages that exhibit the gliding behavior class.

Behaviors are correlated with certain sequence variants

Analyses of trait evolution show a correlation between one behavior and one representative sequence variant. There is very strong evidence (log BF = 11.3268) and a significant association between three gliding species and the sequence variant B clade (OPNGq2): ln(indepen) = -21.9467; ln(depen) = -16.2851; LRT = 11.3232; $P(X^2_{df=4})$ = 0.0231; ($AIC_{indepen}$ = 47.8934; AIC_{depen} = 34.5702) (Table 3-S5). Two species, *Ylistrum japonicum* and *Y. balloti*, represented in sequence variant B are sister to one another while the third species, *Placopecten magellanicus*, is equally related to all remaining gliding species (with the exception of *Adamussium colbecki*). When we compared the

amino acid sequence variant that comprised the B clade to the remaining sequence variants, there was no difference between amino acids at putative spectral tuning sites. However, when we expanded the search to include amino acid positions outside the putative spectral tuning sites, we found three nonsynonymous differences between sequence variants that comprise the B clade identical sequence group and the majority of the remaining OPNGq2 sequence variants at positions V214L, D288N, and S297A (Table 3-6). This pattern was also observed in two other gliding species, *Amusium pleuronectes* and *Amusium papyraceum*, but their representative sequence variants each contain an additional nonsynonymous change at L303S and E247G, respectively, which are not shared by sequence variants that comprise the B clade. As discussed in the above sections, one pattern seen exclusively in gliding species was the presence of a neutral, polar asparagine (N) at position 288, rather than an acidic polar aspartic acid (D). This position is located two amino acids downstream of transmembrane VI, on the extracellular side (Figure 2-1 in (Porath-Krause *et al*., 2016). Analyses of trait evolution revealed no behaviors are correlated with sequence variant represented by clades A, C, D, or E (Table 3-S5). Additionally, no median depth category is significantly correlated with the sequence variants (Table 3-S6).

Discussion

We demonstrate that two G_q-opsin paralogs have different evolutionary trajectories across Pectinidae. Quantifying and comparing rates of mutation in G_q-opsin paralogs is an initial step to explore molecular processes explaining evolutionary patterns

of positive and diversifying selection. Here, we found *opnGq2* is under stronger purifying selection than *opnGq1* caused by an increase in the nonsynonymous substitution rate across the *opnGq1* clade, which led to a higher rate of evolution in *opnGq1* when compared to *opnGq2*. The significant difference in omega between *opnGq1* and *opnGq2* suggests there were different selection pressures for each G_q-opsin paralog, followed by purifying selection to maintain the paralogs. Only one codon position, 246, was identified to be under diversifying selection in *opnGq1*. This position is located inside transmembrane VI, but is not one of the putative spectral tuning positions in squid- G_q-opsin (Sekharan *et al.*, 2012). There are no other positions or transmembrane regions under positive selection in either *opnGq1* or *opnGq2*. We did, however, identify a higher number of codon positions in *opnGq2* under purifying selection than in *opnGq1*. We expect that the significant difference in dN/dS ratios is not the result of positive selection for functional divergence (spectral tuning), but is the result of a stronger force of purifying selection for one G_q-opsin paralog over the other. One mechanism that could explain a stronger force of purifying selection of *opnGq2* compared to *opnGq1* is translational robustness (Drummond *et al.*, 2005). Translational robustness is the ability of a protein to fold correctly even in the presence of missense substitutions. While this mechanism seems a paradox to explain low evolutionary rates, if the protein is misfolded due to mistranslation, the result can be deleterious and result in a high fitness cost, especially in highly expressed genes (Wilke, 2006). Mutations that cause a loss of translational robustness will fail to become fixed, so the gene sequence will remain relatively unchanged over time (Wilke, 2006; Drummond & Wilke, 2008). Selection for translational robustness could explain why the very highly expressed

OPNGq2 (over 100,000 FPKM in eye) (Porath-Krause *et al*., 2016) is evolving at a significantly slower rate than OPNGq1 (around 35 FPKM in eye) (Porath-Krause *et al*., 2016).

We identified eight positions, which are not putative spectral tuning sites that may explain the patterns of nucleotide sequence evolution and amino acid sequence convergence. While unequal rates of evolution is a pattern frequently observed in other duplicated opsin gene sequences and is explained by an increase in the rate of amino acid substitution at sites that form the chromophore-binding pocket (Briscoe, 2001; Spaethe, 2004; Li *et al.*, 2009), we were unable to detect a significant difference in evolutionary rates in those G_q-opsin amino acid positions comprising putative spectral tuning sites. When we focused on those positions that contribute to the nonsynonymous differences between sequence variants, we identified positions that contain at least one radical change in amino acid property groups and may play a role in opsin function, such as spectral tuning.

While each of the eight positions may contribute to the difference in the evolutionary rates, position 288 may best explain the source of selection on the OPNGq2 amino acid sequences. Similar to the gliding behavior class that arose in four independent lineages as a result of strong selection for the gliding behavior (Alejandrino *et al.*, 2011), there is strong selection at this particular amino acid position, 288, to maintain an asparagine rather than aspartic acid. While this position is not considered a putative spectral tuning site (Yokoyama & Radlwimmer, 1998; Takahashi & Ebrey, 2003; Sekharan *et al.*, 2012), nonsynonymous differences at this position could still contribute to spectral tuning for two reasons. One, many systems demonstrate positions

that form the chromophore pocket modulate absorbance spectra, yet there have been studies of vertebrate SWS pigments show that positions that do not form the chromophore pocket can act synergistically to modify absorbance spectra (Wilkie *et al.*, 2000). Two, our description of each position location is an estimate based on homology modeling. Position 288 is supposedly located two amino acid positions downstream of transmembrane VI (see Figure 2-1 in Porath-Krause *et al*. 2016), but this is only an approximation for scallops because our estimation of putative spectral tuning sites are based on the x-ray crystallography of *Todarodes pacificus,* which may not reflect the actual protein domains of scallop G_q-opsins. Position 288 could actually be located inside the transmembrane region or is one of the sites that form the chromophore binding pocket. Because we are unclear whether this particular amino acid position or the subsequent differences among Pectinidae species will affect the λ_{max} value, protein expression and site-directed mutagenesis studies are required to resolve which amino acid differences generate functional differences.

Differences in the rate of evolution between *opnGq1* and *opnGq2* sequences could be dependent on scallop traits that involve light. While we do not measure the λ_{max} values of each G_q-opsin sequence, we assume identical sequence variants have the same λ_{max} values for this study. While we included all sites that are estimated to be involved in *Argopecten irradians* G_q-opsin spectral tuning (Porath-Krause *et al*. 2016), one caveat of this study is the missing a portion of the upstream and downstream region of the G_q-opsin genes, which may reveal the identical sequence variants are not identical because we did not capture all amino acid differences. Additional amino acid changes could result in a broader spectrum of expected λ_{max} values across predominant sequence variants. While

we included all sites that are estimated to be involved in *Argopecten irradians* G_q-opsin spectral tuning (Porath-Krause *et al*., 2016).

Based on a significant correlation between the gliding behavior and identical sequence variant B, behavior, not habitat, appears to be a selective force involved with the evolution of OPNGq2. We detected a correlation between three species that exhibit the gliding behavior and sequence variant B. These species are found at similar depths and may be exposed to similar photic environments. All species that exhibit the gliding behavior with the exception of one, *Adamussium colbecki*, are found at similar depths, so one may expect these species would share the same predominant sequence variant. Nonsynonymous differences between these species prevent this pattern. While there are many examples of environment and opsin showing strong correlations suggesting adaptation to light environment (Wang *et al.*, 2004; Spady, 2005; Yokoyama *et al.*, 2008), there are studies that reveal a lack of significance between environment and spectral tuning (Frentiu *et al.*, 2015). While the number of identical OPNGq1 amino acid sequence variants was quite high (nearly 60% of species tested) compared to the maximum of three (<10% in a group) sequences that are identical to one another in OPNGq2, the absence of a correlation between clade A and behaviors or depth suggests neither trait is the sole cause of the evolution of OPNGq1 sequence variants. So, why do so many scallop species continue to retain OPNGq1 sequences? What is the source of selection driving the variation in sequence variants across OPSGq2? Including positions outside the frequently studied putative spectral tuning sites may reveal promising candidate positions that may explain interesting patterns of opsin evolution. Broadening the potential positions of interest to include sites located outside the well-characterized

putative spectral tuning sites will capture important patterns of change that may have otherwise been lost. While identifying signatures of selection and identifying correlations can provide estimates of putative spectral tuning sites, it is just the first step in addressing questions of adaptation; expressing the proteins *in vitro* is the most robust way to test hypotheses linking spectral tuning sites to specific environments or behaviors

References

Abascal, F., Zardoya, R. & Posada, D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104–2105.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**: 716–723.

Alejandrino, A., Puslednik, L. & Serb, J.M. 2011. Convergent and parallel evolution in life habit of the scallops (Bivalvia: Pectinidae). *BMC Evol. Biol.* **11**: 164.

Asenjo, A.B., Rim, J. & Oprian, D.D. 1994. Molecular determinants of human red/green color discrimination. *Neuron* **12**: 1131-1138.

Baracchi, D., Petrocelli, I., Cusseau, G. & Pizzocaro, L. 2013. Facial markings in the hover wasps: quality signals and familiar recognition cues in two species of Stenogastrinae. *Animal Behaviour* **85**: 203-212.

Barcellos, L.J.G., Ritter, F., Kreutz, L.C., Quevedo, R.M., da Silva, L.B., Bedin, A.C., *et al.* 2007. Whole-body cortisol increases after direct and visual contact with a predator in zebrafish, Danio rerio. *Aquaculture* **272**: 774–778.

Briscoe, A.D. 2001. Functional Diversification of Lepidopteran Opsins Following Gene Duplication. *Mol. Biol. Evol.* **18**: 2270-2279.

Carleton, K.L. & Kocher, T.D. 2001. Cone opsin genes of african cichlid fishes: tuning spectral sensitivity by differential gene expression. *Mol. Biol. Evol.* **18**: 1540–1550.

Delport, W., Poon, A.F.Y., Frost, S.D.W. & Kosakovsky Pond, S.L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**: 2455–2457.

Dijkstra, H. 1991. A contribution to the knowledge of the pectinacean Mollusca (Bivalvia: Propeamussiidae, Entoliidae, Pectinidae) from the Indonesian Archipelago. *Zool. Verh. Leiden* **271**: 1-57.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., *et al*. 2011. *Geneious v5*. Available: http://www.geneious.com.

Drummond, D.A. & Wilke, C.O. 2008. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell* **134**: 341–352.

Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O. & Arnold, F.H. 2005. Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences* **102**: 14338–14343.

Dulai, K.S., Dornum, von, M., Mollon, J.D. & Hunt, D.M. 1999. The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Res.* **9**: 629–638.

Frentiu, F.D., Bernard, G.D., Sison-Mangus, M.P., Brower, A.V.Z. & Briscoe, A.D. 2007. Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. *Mol. Biol. Evol.* **24**: 2016–2028.

Frentiu, F.D., Yuan, F., Savage, W.K., Bernard, G.D., Mullen, S.P. & Briscoe, A.D. 2015. Opsin Clines in Butterflies Suggest Novel Roles for Insect Photopigments. *Mol. Biol. Evol.* **32**: 368–379.

Gu, X. 2001. Maximum-likelihood approach for gene family evolution under functional divergence. *Mol. Biol. Evol.* **18**: 453–464.

Gu, X. & Vander Velden, K. 2002. DIVERGE: phylogeny-based analysis for functional–structural divergence of a protein family. *Bioinformatics*.

Guindon, S. & Gascuel, O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol*. **52**: 696-704.

Gumbert, A. 2000. Color choices by bumble bees (Bombus terrestris): innate preferences and generalization after learning. *Behavioral Ecology and Sociobiology* **48**: 36–43.

Hamilton, P.V. & Koch, K.M. 1996. Orientation toward natural and artificial grassbeds by swimming bay scallops, Argopecten irradians (Lamarck, 1819). *Journal of Experimental Marine Biology and Ecology* **199**: 79–88.

Hasegawa, M., Kishino, H. & Yano, T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol*. **22**: 160-174.

Hoffmann, M., Tripathi, N., Henz, S.R., Lindholm, A.K., Weigel, D., Breden, F., *et al.* 2007. Opsin gene duplication and diversification in the guppy, a model for sexual selection. *Proceedings of the Royal Society B: Biological Sciences* **274**: 33–42.

Hunt, D.M., Carvalho, L.S., Cowing, J.A. & Davies, W.L. 2009. Evolution and spectral tuning of visual pigments in birds and mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**: 2941–2955.

Innan, H. & Kondrashov, F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* **11**: 4.

Jenkins, G.M., Rambaut, A., Pybus, O.G. & Holmes, E.C. 2014. Rates of Molecular Evolution in RNA Viruses: A Quantitative Phylogenetic Analysis. *J Mol Evol* **54**: 156–165.

Jerlov, N.G. 1976. *Jerlov: Marine Optics,* 229.

Jones, D.T., Taylor, W.R. & Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.

Jordan, I.K., Wolf, Y.I. & Koonin, E.V. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol. Biol.* **4**: 22.

Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**: 3059–3066.

Koyanagi, M., Nagata, T., Katoh, K., Yamashita, S. & Tokunaga, F. 2008. Molecular Evolution of Arthropod Color Vision Deduced from Multiple Opsin Genes of Jumping Spiders. *J Mol Evol* **66**: 130–137.

Lemey, P., Kosakovsky Pond, S.L., Drummond, A.J., Pybus, O.G., Shapiro, B., Barroso, H., *et al.* 2007. Synonymous Substitution Rates Predict HIV Disease Progression as a Result of Underlying Replication Dynamics. *PLoS Comput Biol* **3**: e29.

Li, Z., Gan, X. & He, S. 2009. Distinct Evolutionary Patterns Between Two Duplicated Color Vision Genes Within Cyprinid Fishes. *J Mol Evol*. **69**: 346-359.

Light, J. & Hafner, M. 2008. Codivergence in Heteromyid Rodents (Rodentia: Heteromyidae) and Their Sucking Lice of the Genus Fahrenholzia (Phthiraptera: Anoplura). *Systematic Biol.* **57**: 449–465.

Lythgoe, J.N. 1968. Visual pigments and visual range underwater. *Vision Res.* **8**: 997–1011.

Marais, G. & Duret, L. 2001. Synonymous codon usage, accuracy of translation, and gene length in Caenorhabditis elegans. *J. Mol. Evol*. **52**: 275-280.

Morton, B. 2008. The Evolution of Eyes in the Bivalvia: New Insights*. *American Malacological Bulletin* **26**: 35–45.

Murakami, M. & Kouyama, T. 2008. Crystal structure of squid rhodopsin. *Nature* **453**: 363–367.

Nei, M. & Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.

Nei, M. & Kumar, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, USA.

Neitz, M., Neitz, J. & Jacobs, G.H. 1991. Spectral tuning of pigments underlying red-green color vision. *Science* **252**: 971–974.

Ohno, S. 1970. *Evolution by gene duplication.* London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.

Ohta, T. 1994. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. *Genetics* **138**: 1331–1337.

Pagel, M. 1994. Detecting Correlated Evolution on Phylogenies - a General-Method for the Comparative-Analysis of Discrete Characters. *P Roy Soc Lond B Bio* **255**: 37–45.

Pagel, M. & Meade, A. 2006. Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. *The American Naturalist* **167**: 808–825.

Plachetzki, D.C., Degnan, B.M. & Oakley, T.H. 2007. The origins of novel protein interactions during animal opsin evolution. *PLoS ONE* **2**: e1054.

Pond, S.L.K. & Frost, S.D.W. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**: 2531–2533.

Pond, S.L.K. & Frost, S.D.W. 2005b. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol. Biol. Evol.* **22**: 1208-1222.

Porter, M.L., Blasic, J.R., Bok, M.J., Cameron, E.G., Pringle, T., Cronin, T.W., *et al.* 2012. Shedding new light on opsin evolution. *Proc. R. Soc. B.* **279**: 3-14.

Porter, M.L., Speiser, D.I., Zaharoff, A.K., Caldwell, R.L., Cronin, T.W. & Oakley, T.H. 2013. The Evolution of Complexity in the Visual Systems of Stomatopods: Insights from Transcriptomics. *Integrative and Comparative Biology* **53**: 39–49.

Posada, D. 2008. jModelTest: Phylogenetic Model Averaging. *Mol. Biol. Evol.* **25**: 1253–1256.

Puslednik, L. & Serb, J.M. 2008. Molecular phylogenetics of the Pectinidae (Mollusca: Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology. *Mol. Phylogenet. Evol.* **48**: 1178–1188.

Ramirez, M.D., Pairett, A.N., Pankey, M.S., Serb, J.M., Speiser, D.I., Swafford, A.J., et al. 2016. The last common ancestor of bilaterian animals possessed at least 7 opsins. doi:10.1101/052902.

Rennison, D.J., Owens, G.L. & Taylor, J.S. 2012. Opsin gene duplication and divergence in ray-finned fish. *Mol. Phylogenet. Evol.* **62**: 986–1008.

Rosenbaum, D.M., Rasmussen, S.G.F. & Kobilka, B.K. 2009. The structure and function of G-protein-coupled receptors. *Nature* **459**: 356–363.

Sekharan, S., Wei, J.N. & Batista, V.S. 2012. The Active Site of Melanopsin: The Biological Clock Photoreceptor. *J Am Chem Soc* **134**: 19536–19539.

Serb, J.M., Porath-Krause, A.J. & Pairett, A.N. 2013a. Uncovering a Gene Duplication of the Photoreceptive Protein, Opsin, in Scallops (Bivalvia: Pectinidae). *Integrative and Comparative Biology* **53**: 68–77.

Serb, J.M., Porath-Krause, A.J. & Pairett, A.N. 2013b. Uncovering a Gene Duplication of the Photoreceptive Protein, Opsin, in Scallops (Bivalvia: Pectinidae). *Integrative and Comparative Biology*, doi: 10.1093/icb/ict063.

Shi, Y. & Yokoyama, S. 2003. Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates. *Proceedings of the National Academy of Sciences* **100**: 8308–8313.

Smith, A.R. & Carleton, K.L. 2010. Allelic Variation in Malawi Cichlid Opsins: A Tale of Two Genera. *J Mol Evol* **70**: 593–604.

Spady, T.C. 2005. Adaptive Molecular Evolution in the Opsin Genes of Rapidly Speciating Cichlid Species. *Mol. Biol. Evol.* **22**: 1412–1422.

Spaethe, J. 2004. Early Duplication and Functional Diversification of the Opsin Gene Family in Insects. *Mol. Biol. Evol.* **21**: 1583–1594.

Speiser, D.I. & Johnsen, S. 2008. Comparative morphology of the concave mirror eyes of scallops (Pectinoidea). *Amer. Malac. Bull*. **26**: 27-33.

Speiser, D.I., Loew, E.R. & Johnsen, S. 2011. Spectral sensitivity of the concave mirror eyes of scallops: potential influences of habitat, self-screening and longitudinal chromatic aberration. *J. Exp. Biol.* **214**: 422–431.

Takahashi, Y. & Ebrey, T.G. 2003. Molecular basis of spectral tuning in the newt short wavelength sensitive visual pigment. *Biochemistry* **42**: 6025–6034.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725–2729.

Taylor, J.S. & Raes, J. 2004. Duplication and Divergence: The Evolution of New Genes and Old Ideas. *Annu. Rev. Genet.* **38**: 615–643.

Terai, Y., Seehausen, O., Sasaki, T., Takahashi, K., Mizoiri, S., Sugawara, T., *et al.* 2006. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol.* **4**: e433.

WALLER, T.R. 1991. Evolutionary relationships among commercial scallops (Mollusca : Bivalvia : Pectinidae). *Scallops : Biology, Ecology and Aquaculture.* 1–73. Elsevier.

Wang, D., Oakley, T., Mower, J., Shimmin, L.C., Yim, S., Honeycutt, R.L., *et al.* 2004. Molecular evolution of bat color vision genes. *Mol. Biol. Evol.* **21**: 295–302.

Whitmore, A.V. & Bowmaker, J.K. 1989. Seasonal variation in cone sensitivity and short-wave absorbing visual pigments in the rudd Scardinius erythrophthalmus. *J Comp Physiol A* **166**: 103–115.

Wilke, C.O. 2006. Population Genetics of Translational Robustness. *Genetics* **173**: 473–481.

Wilkie, S.E., Robinson, P.R., Cronin, T.W., Poopalasundaram, S., Bowmaker, J.K. & HUNT, D.M. 2000. Spectral Tuning of Avian Violet- and Ultraviolet-Sensitive Visual Pigments. *Biochemistry* **39**: 7895–7901.

Yang, L. & Gaut, B.S. 2011a. Factors that contribute to variation in evolutionary rate among Arabidopsis genes. *Mol. Biol. Evol.* **28**: 2359–2369.

Yang, L. & Gaut, B.S. 2011b. Factors that contribute to variation in evolutionary rate among Arabidopsis genes. *Mol. Biol. Evol.* **28**: 2359–2369.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**: 568-573.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586–1591.

Yokoyama, S. 2000. Molecular evolution of vertebrate visual pigments. *Prog Retin Eye Res* **19**: 385–419.

Yokoyama, S. & Radlwimmer, F.B. 1998. The “five-sites” rule and the evolution of red and green color vision in mammals. *Mol. Biol. Evol.* **15**: 560–567.

Yokoyama, S., Tada, T., Zhang, H. & Britt, L. 2008. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **105**: 13480–13485.

Yuan, Q., Metterville, D., Briscoe, A.D. & Reppert, S.M. 2007. Insect Cryptochromes: Gene Duplication and Loss Define Diverse Ways to Construct Insect Circadian Clocks. *Mol. Biol. Evol.* **24**: 948–955.

Zuckerkandl, E. 1976. Evolutionary processes and evolutionary noise at the molecular level. *J Mol Evol*. **7**: 269-311.

Table 3-1 Primers used to amplify scallop G_q-opsin paralogs

Gene	Primer	Sequence (5' → 3'): forward and reverse	Annealing Temperature (°C)	Primer Concentration (mM)	Product Size (bp)
opnGq1	530F 1076R	GATCGTCGTCGTTT GGGTTTGGCAGTTG AGGAAACAAGGCG	STEP UP[a]: 54, 57	10	540-588
opnGq2	RhoF2 1100R	CATGACGAGACGC AAGGTCCACCGAT GGCTGAGAGCATA TACCACTGG	STEP UP[a]: 52, 55	10	521
opnGq2	355F 1100R	ATCTGGCGGTCAGT GACCTCATCTTCTC GATGGCTGAGAGC ATATACCACTGG	STEP UP[a]: 50, 55 STEP UP[a]: 55, 60	10	541-973
opnGq2	GigRho1R[b] 1100R [b]	GCCTTCATACTCCT GGTCATAGGATGG CTGAGAGCATATAC CACTGG	STEP UP[a]: 50, 55	12	493

[a] STEP UP thermal cycle profile: 96°C 1min; (96°C 1min, first annealing temp for 2min, 72°C 4min) 15x; (96°C 1min, second annealing temp for 2min, 72°C 4min) 15x; final extension (72°C 30min)

[b] Used to extend opsin sequence in *Crassadoma gigantea*

Table 3-2 Species included in analyses

Species	NCBI accession number	Behavior	Median depth	Presence of opnGq1 (1) or opnGq2 (2)
Ctenoides annulatus	XXX	_	146.5	1, 2
Bractechlamys antillarum	XXX	Byssal attaching	65	1, 2
Ylistrum balloti	XXX	Gliding	45	1, 2
Leptopecten bavayi	XXX	Byssal attaching	25	1, 2
Spathochlamys benedicti	XXX	Byssal attaching	65	1, 2
Mimachlamys cloacata	XXX	Free living	68	1, 2
Adamussium colbecki	XXX	Gliding	70	1, 2
Propeamussium dalli	XXX	_	892	1, 2
Veprichlamys empressae	XXX	Byssal attaching	195	1, 2
Pecten fumatus	XXX	Recessing	46.58	1, 2
Crassodoma gigantea	XXX	Nestling	45.5	1
Chlamys hastata	XXX	Byssal attaching	18	1, 2
Spondylus ictericus	XXX	_	70.5	1, 2
Argopecten irradians	XXX	Free living	5	1, 2
Ylistrum japonicum	XXX	Gliding	45	1, 2
Laevichlamys lemniscata	XXX	Byssal attaching	15.5	1, 2
Placopecten magellanicus	XXX	Gliding	55	1, 2
Semipallium marybellae	XXX	Byssal attaching	19.5	1, 2
Pecten maximus	XXX	Recessing	57.5	1, 2
Talochlamys multistriata	XXX	Byssal attaching	55	1, 2
Aequipecten opercularis	XXX	Free living	32.5	1, 2
Caribachlamys ornata	XXX	Byssal attaching	10	1, 2
Amusium papyraceum	XXX	Gliding	48.5	1, 2
Amusium pleuronectes	XXX	Gliding	47	1, 2
Decatopecten plica	XXX	Free living	27	1, 2
Talochlamys pusio	XXX	Cementing	50	1, 2
Chlamys rubida	XXX	Byssal attaching	48	1, 2
Mimachlamys sanguinea	XXX	Byssal attaching	4.5	1, 2
Pseudamussium septemradiatum	XXX	Free living	65	1, 2
Pedum spondyloideum	XXX	Nestling	12.75	1, 2
Cryptopecten vesiculosus	XXX	Free living	68.5	1, 2
Euvola vogdesi	XXX	Recessing	50	1, 2
Mizuhopecten yessoensis	XXX	Recessing	27.5	1, 2
Euvola ziczac	XXX	Recessing	29	1, 2

Table 3-3 Rates of evolution for each test

Test	Model	Ln	*opnGq1* omega	*opnGq2* omega	Outgroup omega
H0	0	-5531.856	0.071	0.071	0.071
H1	2	-5527.848	0.0728	0.0728	0.0728
H2	2	-5524.41	0.1114	0.0568	0.1114
H3	2	-5520.413	0.1192	0.0568	0.0025
H4	2	-5527.841	0.0722	0.0722	0.0722
H5	2	-5522.824	0.1181	0.0556	0.0556
H6	2	-5520.413	0.1192	0.0568	0.0025

Table 3-4 Testing for variable rates of evolution between *opnGq1* and *opnGq2*

Hypothesis and test	LRT	LRT Value	$P(X^2_{df=1})$
Hyp[A]: T0 vs. T1	2(-5527.85--5531.86)	8.016	0.0046
Hyp[B]: T0 vs. T2	2(-5524.41--5531.86)	14.892	0.0001
Hyp[C]: T2 vs. T3	2(-5520.41--5524.41)	7.994	0.0047
Hyp[D]: T0 vs. T4	2(-5527.85--5531.86)	8.01	0.0047
Hyp[E]: T0 vs. T5	2(-5522.82--5531.86)	18.064	2.14E-05
Hyp[F]: T5 vs. T6	2(-5520.41--5522.84)	4.822	0.0281

[A] The mutation rate of *opnGq2* has changed relative to *opnGq1*.

[B] A burst of positive selection for functional divergence occurred following the duplication event that gave rise to *opnGq2*.

[C] There was a long-term shift in selective constraints following the duplication event that gave rise to *opnGq2*.

[D] The mutation rate of *opnGq1* has increased relative to *opnGq2*.

[E] A burst of positive selection for functional divergence occurred following the duplication event that gave rise to *opnGq1*.

[F] A long-term shift in selective constraints followed the duplication event that gave rise to *opnGq1*.

Table 3-5 Putative spectral tuning sites

Position Number	[e]177	[e]185	[f]201	[f]204	[f]209	[f]301	[f]303	[f]306[d]	[f]308[d]	[f]309
OPNGq1 | OPNGq2	Y | Y[c]	S | S[c]	[b]I | I[c]	L | M	[b]F | F[c]	M | V[c]	[b]L | L	S | A[ac]	[b]S | A[a]	M | M
C. annulatus	. | .	. | .	. | .	. | .	. | .	. | .	. | S[a]	. | .	. | .	. | .
B. antillarum	. | .	. | .	. | .	. | .	. | .	. | .	. | .	T | .	. | .	. | .
Y. balloti	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
L. bavayi	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
S. benedicti	. | .	. | P[a]	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
M. cloacata	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
A. colbecki	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
P. dalli	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
V. empressae	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
P. fumatus	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
C. gigantea	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
C. hastata	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
S. ictericus	. | H[a]	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
A. irradians	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
Y. japonicum	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
L. lemniscata	. | .	. | .	. | M	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
P. magellanicus	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
S. marybellae	. | .	. | .	. | M	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | I
P. maximus	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
T. multistriata	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
A. opercularis	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
C. ornata	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
A. papyraceum	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
A. pleuronectes	. | .	. | .	. | .	. | .	. | .	. | .	. | S[a]	. | .	. | .	. | .	. | .
D. plica	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
T. pusio	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
C. rubida	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
M. sanguinea	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
P. septemradiatum	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
P. spondyloideum	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
C. vesiculosus	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
E. vogdesi	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .
M. yessoensis	. | .	. | .	. | .	. | .	. | Y[a]	. | .	. | .	. | S[a]	. | .	. | .	. | .
E. ziczac	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .	. | .

[a] Radical difference in the amino acid group, [b] Purifying selection identified in opnGq1

[c] Purifying selection identified in opnGq2, [d] Positions containing Type II Divergence

[e] Position in the extra-cellular region of the protein, [f] Position located in the transmembrane region of the protein

Table 3-6 Positions of interest

Position Number	[e]195[d]	[f]214[d]	[f]217[d]	[f]221[d]	[f]247[d]	[f]265[d]	[e]288[d]	[f]297[d]
OPNGq1 | OPNGq2	A | T[a]	I | V	G | I[c]	V | I[c]	[b]R | E[ac]	T | T[c]	[b]P | D[ac]	S | S
C. annulatus	. | L[a]	. | .	. | .	. | V	. | .	. | .	. | .	. | A[a]
B. antillarum	P | .	L | M	[a]S | .	I | .	. | .	[a]A | .	. | .	[a]A | .
Y. balloti	. | .	. | L	. | V	. | .	. | .	. | .	. | N[a]	. | A[a]
L. bavayi	. | L[a]	. | .	. | .	. | .	. | .	. | M[a]	. | .	. | .
S. benedicti	. | L[a]	. | .	. | .	. | .	. | .	. | M[a]	. | .	. | .
M. cloacata	. | L[a]	. | .	. | .	. | .	. | .	. | M[a]	. | .	. | .
A. colbecki	. | .	. | .	. | V	. | .	. | D	. | M[a]	. | N[a]	. | .
P. dalli	. | .	. | F	. | .	. | .	. | D	. | .	. | .	. | .
V. empressae	. | L[a]	. | .	. | V	. | V	. | .	. | .	. | .	. | .
P. fumatus	. | L[a]	. | .	. | .	. | .	. | .	. | M[a]	. | .	. | .
C. gigantea	. | _	V | _	[a]C | _	I | _	. | _	. | _	. | _	. | _
C. hastata	. | V[a]	. | .	[a]C | .	. | .	. | D	. | .	. | .	. | .
S. ictericus	. | .	. | F	. | .	. | .	. | D	. | .	. | .	. | .
A. irradians	. | .	. | G	. | .	. | .	. | D	. | .	. | .	. | .
Y. japonicum	. | .	. | L	. | V	. | .	. | .	. | .	. | N[a]	. | A[a]
L. lemniscata	. | I[a]	V | L	[a]C | .	I | .	. | D	. | .	. | .	. | .
P. magellanicus	. | .	. | L	. | V	. | .	. | .	. | .	. | N[a]	. | A[a]
S. marybellae	M | .	. | C[a]	[a]C | .	. | .	. | D	. | .	. | .	. | .
P. maximus	. | L[a]	. | .	. | .	. | .	. | .	. | M[a]	. | .	. | .
T. multistriata	. | .	V | F	[a]C | .	. | N[a]	. | D	. | .	. | .	. | .
A. opercularis	. | .	. | .	. | .	. | V	. | .	. | .	. | .	. | .
C. ornata	. | .	. | M	. | .	. | .	. | D	. | .	. | .	. | .
A. papyraceum	. | .	. | L	. | V	. | .	. | G[a]	. | .	. | N[a]	. | A[a]
A. pleuronectes	. | .	. | L	. | V	. | .	. | .	. | .	. | N[a]	. | A[a]
D. plica	. | I[a]	. | M	. | V	. | .	. | .	. | .	. | .	. | .
T. pusio	. | I[a]	V | G	[a]C | V	I | .	. | D	. | .	. | .	. | .
C. rubida	. | .	. | G	[a]C | .	. | .	. | D	. | .	. | .	. | .
M. sanguinea	. | .	V | F	[a]C | .	. | .	. | D	. | .	. | .	. | .
P. septemradiatum	. | V[a]	. | L	. | .	. | .	. | D	. | M[a]	. | .	. | .
P. spondyloideum	. | V[a]	. | L	. | .	. | .	. | D	. | M[a]	. | .	. | .
C. vesiculosus	. | L[a]	. | .	. | .	. | V	. | .	. | .	. | .	. | A[a]
E. vogdesi	. | L[a]	. | .	. | V	. | V	. | .	. | .	. | .	. | .
M. yessoensis	. | V[a]	. | M	. | .	. | .	. | D	. | .	. | .	. | .
E. ziczac	. | L[a]	. | .	. | V	. | V	. | .	. | .	. | .	. | A[a]

[a] Radical difference in the amino acid group, [b] Purifying selection identified in opnGq1

[c] Purifying selection identified in opnGq2, [d] Positions containing Type II Divergence

[e] Position in the extra-cellular region of the protein, [f] Position located in the transmembrane region of the protein

Table 3-S1 Duplicate, Depth, Behavior information

Species	Reference name	Catalogue or reference number	Location of collection	opnGq1 Sequence Variant clade	opnGq2 Sequence Variant clade	opnGq1 Identical Sequence Variant Yes(1), No(0)	opnGq2 Identical Sequence Variant Yes(1), No(0)	Depth category	Depth reference 1	Depth reference 2
Ctenoides annulatus	annulatus	UF322180	New Guinea	A	F	1	0	Very deep	Encylopedia of Life, 2015, http://eol.org	
Bractechlamys antillarum	antillarum	1	not given	F	F	0	0	Medium	Ghys, 2015, http://www.pectensite.com/	
Ylistrum balloti	balloti	1	Australia	A	B	1	1	Medium	Speiser & Johnsen, 2008$_C$	Himmelman et al., 2009$_E$
Leptopecten bavayi	bavayi	UF371875	Panama	F	D	0	1	Medium	Arne Ghys, 2015, http://www.pectensite.com/	
Spathochlamys benedicti	benedicti	UF369432	Fiji	A	F	1	0	Medium	Waller, 1993$_F$	
Mimachlamys cloacata	cloacata	UF309990	Taiwan	F	F	0	0	Deep	Arne Ghys, 2015, http://www.pectensite.com/	
Adamussium colbecki	colbecki	3	Antarctica	A	F	1	0	Deep	Heilmayer et al., 2005$_A$	Arne Ghys, 2015, http://www.pectensite.com/
Propeamussium dalli	dalli	UF289879	USA	A	F	1	0	Very deep	Bill Frank, 2015, http://www.jaxshells.org	Encylopedia of Life, 2015, http://eol.org
Veprichlamys empressae	empressae	HPC1578	Japan	A	A	1	1	Very deep	Arne Ghys, 2015, http://www.pectensite.com/	
Pecten fumatus	fumatus	1	Australia	A	D	1	1	Medium	Arne Ghys, 2015, http://www.pectensite.com/	
Crassodoma gigantea	gigantea	1	USA	F	F	0	0	Medium	Minchin, 2003$_B$	Arne Ghys, 2015, http://www.pectensite.com/
Chlamys hastata	hastata	B001	USA	F	F	0	0	Shallow	Speiser & Johnsen, 2008$_C$	Arne Ghys, 2015, http://www.pectensite.com/
Spondylus ictericus	ictericus	UF367487	USA	A	F	1	0	Deep	Encylopedia of Life, 2015, http://eol.org	
Argopecten irradians	irradians	1	USA*	A	C	1	1	Shallow	Minchin, 2003$_B$	
Ylistrum japonicum	japonicum	1	Canada	A	B	1	1	Medium	Minchin, 2003$_B$	Arne Ghys, 2015, http://www.pectensite.com/
Laevichlamys lemniscata	lemniscata01	2	Japan	F	F	0	0	Shallow	Arne Ghys, 2015, http://www.pectensite.com/	
Placopecten magellanicus	magellanicus	1	USA	A	B	1	1	Medium	Minchin, 2003$_B$	
Semipallium marybellae	marybellae	UF287521	Mariana Islands	F	F	0	0	Shallow	Arne Ghys, 2015, http://www.pectensite.com/	
Pecten maximus	maximus	1	Scotland	A	D	1	1	Medium	Minchin, 2003$_B$	Arne Ghys, 2015, http://www.pectensite.com/
Talochlamys multistriata	multistriata01	1	Spain	F	F	0	0	Medium	Waller, 1993$_F$	Arne Ghys, 2015, http://www.pectensite.com/
Aequipecten opercularis	opercularis	2	Scotland	A	F	1	0	Medium	Minchin, 2003$_B$	Waller, 1993$_F$
Caribachlamys ornata	ornata01	2	Peurto Rico	A	F	1	0	Shallow	Waller, 1993$_F$	
Amusium papyraceum	papyraceum	1	not given	A	F	1	0	Medium	Minchin, 2003$_B$	
Amusium pleuronectes	pleuronectes	3	Thailand	A	F	1	0	Medium	Minchin, 2003$_B$	Arne Ghys, 2015, http://www.pectensite.com/
Decatopecten plica	plica	1	Japan	F	F	0	0	Medium	Arne Ghys, 2015, http://www.pectensite.com/	
Talochlamys pusio	pusio	2	Spain	F	F	0	0	Medium	Waller, 1993$_F$	
Chlamys rubida	rubida	B000	USA	F	C	0	1	Medium	Minchin, 2003$_B$	Arne Ghys, 2015, http://www.pectensite.com/
Mimachlamys sanguinea	sanguinea01	2	not given	F	F	0	0	Shallow	Minchin, 2003$_B$	
Pseudamussium septemradiatum	septemradiatus	2	Scotland	A	F	1	0	Medium	Arne Ghys, 2015, http://www.pectensite.com/	
Pedum spondyloideum	spondyloideum	UF343587	Mariana Islands	F	F	0	0	Shallow	Scaps, 2011$_D$	Arne Ghys, 2015, http://www.pectensite.com/
Cryptopecten vesiculosus	vesiculosus	1	Japan	A	F	1	0	Deep	Arne Ghys, 2015, http://www.pectensite.com/	
Euvola vogdesi	vogdesi	1	Mexico	A	A	1	1	Medium	Used *Euvola perula* Arne Ghys, 2015, http://www.pectensite.com/	
Mizuhopecten yessoensis	yessoensis	2	Japan	A	F	1	0	Medium	Minchin, 2003$_B$	
Euvola ziczac	ziczac	1	Bermuda	F	F	0	0	Medium	Minchin, 2003$_B$	

Table 3-S2 Codon-based test of diversifying selection
Above the diagonal, pairwise comparison values representing *opnGq2* clade. Below the diagonal, pairwise comparison values representing *opnGq1* clade. No pairwise comparison values were significantly under diversifying selection.

Species	*Bractechlamys antillarum*	*Ctenoides annulatus*	*Ylistrum balloti*	*Leptopecten bavayi*	*Spathochlamys benedicti*	*Mimachlamys cloacata*	*Adamussium colbecki*	*Propeamussium dalli*	*Veprichlamys empressae*	*Pecten fumatus*	*Crassodoma gigantea*	*Chlamys hastata*	*Spondylus ictericus*	*Argopecten irradians*	*Ylistrum japonicum*	*Laevichlamys lemniscata*	*Placopecten magellanicus*	*Semipallium marybellae*	*Pecten maximus*	*Talochlamys multistriata*	*Aequipecten opercularis*	*Caribachlamys ornata*	*Amusium papyraceum*	*Amusium pleuronectes*	*Decatopecten plica*	*Talochlamys pusio*	*Chlamys rubida*	*Mimachlamys sanguinea*	*Pseudamussium septemradiatum*	*Pedum spondyloideum*	*Cryptopecten vesiculosus*	*Euvola vogdesi*	*Mizuhopecten yessoensis*	*Euvola zizac*
Bractechlamys antillarum		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ctenoides annulatus	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.430	1.000	1.000	1.000
Ylistrum balloti	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.080	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Leptopecten bavayi	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Spathochlamys benedicti	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Mimachlamys cloacata	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Adamussium colbecki	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Propeamussium dalli	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.080	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Veprichlamys empressae	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Pecten fumatus	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		NA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Crassodoma gigantea	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chlamys hastata	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Spondylus ictericus	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.080	1.000	1.000	1.000	1.000	1.000
Argopecten irradians	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ylistrum japonicum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Laevichlamys lemniscata	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Placopecten magellanicus	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Semipallium marybellae	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Pecten maximus	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Talochlamys multistriata	1.000	1.000	1.000	1.000	1.000	0.159	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Aequipecten opercularis	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Caribachlamys ornata	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Amusium papyraceum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Amusium pleuronectes	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Decatopecten plica	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Talochlamys pusio	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Chlamys rubida	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Mimachlamys sanguinea	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000	1.000
Pseudamussium septemradiatum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000
Pedum spondyloideum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.393	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000	1.000
Cryptopecten vesiculosus	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000	1.000
Euvola vogdesi	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000	1.000
Mizuhopecten yessoensis	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000
Euvola zizac	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		1.000

Table 3-S3 Codon-based test of purifying selection

Above the diagonal, pairwise comparison values representing *opnGq2* clade. Below the diagonal, pairwise comparison values representing *opnGq1* clade. Highlighted cells represent those species pairwise comparison values that were NOT significantly under purifying selection. All other cells represent pairwise comparison values of p less than 0.05 and were considered significant.

Species	*Bractechlamys antillarum*	*Ylistrum balloti*	*Leptopecten bavayi*	*Spathochlamys benedicti*	*Mimachlamys cloacata*	*Adamussium colbecki*	*Ctenoides annulatus*	*Talochlamys pusio*	*Veprichlamys empressae*	*Pecten fumatus*	*Chlamys hastata*	*Argopecten irradians*	*Ylistrum japonicum*	*Laevichlamys lemniscata*	*Placopecten magellanicus*	*Semipallium marybellae*	*Pecten maximus*	*Talochlamys multistriata*	*Crassodoma gigantea*	*Aequipecten opercularis*	*Caribachlamys ornata*	*Amusium papyraceum*	*Amusium pleuronectes*	*Decatopecten plica*	*Propeamussium dalli*	*Chlamys rubida*	*Mimachlamys sanguinea*	*Pseudamussium septemradiatum*	*Pedum spondyloideum*	*Spondylus ictericus*	*Cryptopecten vesiculosus*	*Euvola vogdesi*	*Mizuhopecten yessoensis*	*Euvola ziczac*
Bractechlamys antillarum		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ylistrum balloti	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	1.000	0.000	0.000	0.000	NA	0.000	0.000	0.253	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Leptopecten bavayi	0.000	0.262		0.252	0.121	0.000	0.000	0.000	0.000	0.160	0.000	0.000	0.000	0.000	0.000	0.000	0.081	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Spathochlamys benedicti	0.000	1.000	0.262		0.356	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.252	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mimachlamys cloacata	0.000	1.000	0.377	1.000		0.000	0.000	0.000	0.000	0.252	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adamussium colbecki	0.000	0.160	1.000	0.160	0.261		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ctenoides annulatus	0.000	1.000	0.262	1.000	1.000	0.160		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
Talochlamys pusio	0.000	0.001	0.001	0.001	0.001	0.001	0.001		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veprichlamys empressae	0.000	0.160	0.126	0.160	0.261	0.081	0.160	0.001		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.064
Pecten fumatus	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160		0.000	0.000	0.000	0.000	0.000	0.000	0.160	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Chlamys hastata	0.000	0.006	0.005	0.006	0.008	0.004	0.006	0.000	0.004	0.006		0.000	0.000	0.001	0.000	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.081	0.000
Argopecten irradians	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006		0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ylistrum japonicum	0.000	0.160	0.126	0.160	0.261	0.081	0.160	0.001	0.081	0.160	0.004	0.160		0.000	1.000	0.000	0.000	0.000	NA	0.000	0.000	0.253	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Laevichlamys lemniscata	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001		0.000	0.000	0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000
Placopecten magellanicus	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001		0.000	0.000	0.000	NA	0.000	0.000	0.253	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Semipallium marybellae	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pecten maximus	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000		0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Talochlamys multistriata	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		NA	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.249	0.000	0.000	0.458	0.000	0.000	0.000	0.000
Crassodoma gigantea	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.261	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.000	0.001	0.000		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Aequipecten opercularis	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Caribachlamys ornata	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001	1.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Amusium papyraceum	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001	1.000	1.000		0.469	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Amusium pleuronectes	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001	1.000	1.000	1.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Decatopecten plica	0.000	0.136	0.103	0.136	0.182	0.075	0.136	0.002	0.075	0.136	0.008	0.136	0.075	0.002	0.136	0.000	0.136	0.000	0.002	0.136	0.136	0.136	0.136		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Propeamussium dalli	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001	1.000	1.000	1.000	1.000	0.136		0.000	0.495	0.000	0.000	1.000	0.000	0.000	0.000	0.000
Chlamys rubida	0.000	0.024	0.020	0.024	0.033	0.014	0.024	0.001	0.014	0.024	0.004	0.024	0.014	0.001	0.024	0.000	0.024	0.000	0.000	0.024	0.024	0.024	0.024	0.036	0.024		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mimachlamys sanguinea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	1.000	0.000	0.000	0.000	0.000
Pseudamussium septemradiatum	0.000	0.160	0.126	0.160	0.261	0.081	0.160	0.001	0.081	0.160	0.004	0.160	0.081	0.001	0.160	0.000	0.160	0.000	0.001	0.160	0.160	0.160	0.160	0.075	0.160	0.014	0.000		0.122	0.000	0.000	0.000	0.000	0.000
Pedum spondyloideum	0.000	0.496	0.331	0.496	1.000	0.253	0.496	0.001	0.253	0.496	0.008	0.496	0.253	0.001	0.496	0.000	0.496	0.000	0.001	0.496	0.496	0.496	0.496	0.137	0.496	0.035	0.000	0.253		0.000	0.000	0.000	0.000	0.000
Spondylus ictericus	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001	1.000	1.000	1.000	1.000	0.136	1.000	0.024	0.000	0.160	0.496		0.000	0.000	0.000	0.000
Cryptopecten vesiculosus	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001	1.000	1.000	1.000	1.000	0.136	1.000	0.024	0.000	0.160	0.496	1.000		0.000	0.000	0.000
Euvola vogdesi	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001	1.000	1.000	1.000	1.000	0.136	1.000	0.024	0.000	0.160	0.496	1.000	1.000		0.000	0.000
Mizuhopecten yessoensis	0.000	1.000	0.262	1.000	1.000	0.160	1.000	0.001	0.160	1.000	0.006	1.000	0.160	0.001	1.000	0.000	1.000	0.000	0.001	1.000	1.000	1.000	1.000	0.136	1.000	0.024	0.000	0.160	0.496	1.000	1.000	1.000		0.000
Euvola ziczac	0.000	0.261	0.183	0.261	0.375	0.126	0.261	0.001	0.126	0.261	0.008	0.261	0.126	0.001	0.261	0.000	0.261	0.000	0.001	0.261	0.261	0.261	0.261	0.102	0.261	0.019	0.000	0.126	0.329	0.261	0.261	0.261	0.261	

Table 3-S4 Branch-site method of site-specific selection

	Null Model Lnl	model A Lnl	Back-ground branch ω	Fore-ground branch ω	Sites under positive selection
opnGq1 clade (as foreground)	-4834.09	-4834.09	0.0621	1	157*, 158, 165*, 196, 199, 203*, 206, 211*, 235*, 239, 240*, 246*, 290
opnGq2 clade (as foreground)	-4860.86	-4860.86	0.0706	1	39, 95, 140

* Significant

Table 3-S5 Relationship between behavior and sequence variant

Bolded values represent p values for the likelihood ratio test with $P(X^2_{df=4})$. Yellow cell represents a significant p-value of less than 0.05. Red cell represents a log bayes factor of very strong evidence of correlated evolution.

opnGq1	Behavior A	LRT | Log BF	Behavior B	LRT | Log BF	Behavior C	LRT | Log BF	Behavior D	LRT | Log BF	Behavior E	LRT | Log BF	Behavior F	LRT | Log BF
ML Indep	-29.2313	0.9444	-40.1205	8.6458	-34.2743	1.5572	-33.6798	6.3380	-35.4779	6.6452	-29.0614	6.7684
ML Depen	-28.7591	**0.9181**	-35.7976	**0.0706**	-33.4957	**0.8165**	-30.5108	**0.1753**	-32.1553	**0.1559**	-25.6772	**0.1486**
MCMC Indep	-30.7175	2.6098	-45.0913	5.0614	-37.5749	1.4204	-35.8834	0.3828	-35.7975	1.7758	-29.6728	3.5358
MCMC Depen	-32.0224		-42.5606		-38.2851		-35.6920		-36.6854		-31.4407	

opnGq2	Behavior A	LRT | Log BF	Behavior B	LRT | Log BF	Behavior C	LRT | Log BF	Behavior D	LRT | Log BF	Behavior E	LRT | Log BF	Behavior F	LRT | Log BF
ML Indep Clade B	-14.0283	0.4014	-26.5893	2.7350	-20.7431	4.5608	-18.4536	0.6924	-21.9467	11.3232	-14.0252	0.7766
ML Depen Clade B	-13.8276	**0.9824**	-25.2218	**0.6031**	-18.4627	**0.3354**	-18.1074	**0.9523**	-16.2851	**0.0232**	-13.6369	**0.9416**
MCMC Indep Clade B	-16.6220	6.0658	-32.0033	1.0292	-26.0002	0.4348	-24.1283	1.0792	-26.2638	11.3268	-16.9042	3.7172
MCMC Depen Clade B	-19.6549		-31.4887		-26.2176		-24.6679		-20.6004		-18.7628	
ML Indep Clade C	-13.8029	0.2556	-26.3871	0.1256	-20.5409	0.8138	-18.2514	1.4092	-21.7445	1.5242	-13.6330	0.2426
ML Depen Clade C	-13.6751	**0.9925**	-26.3243	**0.9981**	-20.1340	**0.9366**	-17.5468	**0.8426**	-20.9824	**0.8223**	-13.5117	**0.9932**
MCMC Indep Clade C	-15.0371	0.4894	-29.3301	0.4890	-23.4302	2.9404	-22.3159	2.6852	-22.4222	4.3909	-14.6724	2.0402
MCMC Depen Clade C	-14.7924		-29.0856		-24.9004		-20.9733		-24.6176		-15.6925	
ML Indep Clade D	-14.8922	0.4268	-27.4764	0.9306	-21.6302	4.4794	-19.3407	3.8840	-22.8339	1.7498	-14.9154	0.7548
ML Depen Clade D	-14.6788	**0.9802**	-27.0111	**0.9201**	-19.3905	**0.3450**	-17.3987	**0.4219**	-21.9590	**0.7817**	-14.5380	**0.9444**
MCMC Indep Clade D	-16.7921	1.5666	-31.7163	2.5156	-26.1989	3.3096	-23.2377	4.3712	-26.5413	2.0114	-16.7413	3.5124
MCMC Depen Clade D	-17.5754		-32.9741		-27.8537		-21.0521		-25.5356		-18.4975	
ML Indep Clade E	-13.7931	0.2518	-26.3546	0.1470	-20.5376	0.9890	-18.2500	0.7592	-21.7120	0.7746	-13.6005	0.2652
ML Depen Clade E	-13.6672	**0.9927**	-26.4281	**0.9974**	-20.0431	**0.9115**	-17.8704	**0.9438**	-21.3247	**0.9418**	-13.4679	**0.9919**
MCMC Indep Clade E	-15.9003	2.9500	-29.5995	1.0856	-22.8188	1.0574	-20.5909	6.0906	-23.3895	2.3028	-15.6671	3.1172
MCMC Depen Clade E	-14.4253		-30.1423		-22.2901		-23.6362		-24.5409		-14.1085	
ML Indep Clade F	-26.1860	1.5862	-38.7702	0.6664	-32.9240	0.5692	-30.6345	2.2180	-34.4465	4.8380	-26.0161	1.5496
ML Depen Clade F	-25.3929	**0.8113**	-39.1034	**0.9554**	-33.2086	**0.9664**	-29.5255	**0.6957**	-32.0275	**0.3043**	-25.2413	**0.8178**
MCMC Indep Clade F	-27.2445	1.2480	-43.2083	0.0856	-35.7960	0.7212	-35.1011	5.3244	-36.0963	1.0038	-27.3905	2.7540
MCMC Depen Clade F	-27.8685		-43.1655		-36.1566		-32.4389		-35.5944		-28.7675	

Table 3-S6 Relationship between depth and sequence variant
Bolded values represent p values for the likelihood ratio test with $P(X^2_{df=4})$.

opnGq1	Shallow	LRT | Log BF	Medium	LRT | Log BF	Deep	LRT | Log BF	Very deep	LRT | Log BF
ML Indep	-38.7892	7.3720	-43.3624	1.0368	-32.2687	0.6636	-29.2313	3.6916
ML Depen	-35.1032	**0.1175**	-42.8440	**0.9042**	-31.9369	**0.9557**	-27.3855	**0.4493**
MCMC Indep	-39.5182	0.6368	-45.5172	0.0352	-34.5183	0.0394	-32.1168	2.0488
MCMC Depen	-39.8366		-45.4996		-34.5380		-33.1412	

opnGq2	Shallow	LRT | Log BF	Medium	LRT | Log BF	Deep	LRT | Log BF	Very deep	LRT | Log BF
ML Indep Clade B	-23.5630	1.1856	-28.1362	0.7488	-18.7375	1.7092	-14.0051	3.8752
ML Depen Clade B	-22.9702	**0.8805**	-27.7618	**0.9452**	-17.8829	**0.7890**	-15.9427	**0.4232**
MCMC Indep Clade B	-27.9419	0.0994	-32.7049	0.8836	-21.3675	2.2950	-19.6389	1.9146
MCMC Depen Clade B	-27.8922		-33.1467		-22.5150		-20.5962	
ML Indep Clade C	-23.3606	0.9812	-27.9340	0.6110	-18.0381	0.5186	-13.8029	2.6120
ML Depen Clade C	-22.8700	**0.9126**	-27.6285	**0.9618**	-17.7788	**0.9717**	-12.4969	**0.6247**
MCMC Indep Clade C	-25.6692	1.8248	-29.5504	0.1472	-18.9324	7.5722	-17.0010	2.5900
MCMC Depen Clade C	-26.5816		-29.6240		-22.7185		-15.7060	
ML Indep Clade D	-24.4501	1.2126	-29.0234	0.7428	-19.6245	1.2488	-16.3434	3.5116
ML Depen Clade D	-23.8438	**0.8760**	-28.6520	**0.9459**	-19.0001	**0.8700**	-14.5876	**0.4761**
MCMC Indep Clade D	-26.4097	5.9168	-31.8360	1.8286	-21.3182	6.5874	-19.4670	3.1008
MCMC Depen Clade D	-29.3681		-30.9217		-24.6119		-21.0174	
ML Indep Clade E	-23.3283	2.3682	-27.9015	1.8176	-18.0056	0.4788	-13.7704	0.6352
ML Depen Clade E	-22.1442	**0.6684**	-26.9927	**0.7693**	-18.2450	**0.9755**	-13.4528	**0.9591**
MCMC Indep Clade E	-24.3654	0.0148	-29.8312	0.5072	-18.8129	0.7290	-16.7868	3.8754
MCMC Depen Clade E	-24.3580		-29.5776		-19.1774		-18.7245	
ML Indep Clade F	-35.8904	1.4112	-40.3172	2.7430	-30.4213	2.0114	-26.1861	1.1820
ML Depen Clade F	-35.1848	**0.8422**	-38.9457	**0.6017**	-29.4156	**0.7337**	-26.7771	**0.1175**
MCMC Indep Clade F	-36.8442	1.3978	-42.8098	0.9714	-31.5312	0.2154	-29.8948	1.5814
MCMC Depen Clade F	-36.1453		-43.2955		-31.6389		-30.6855	

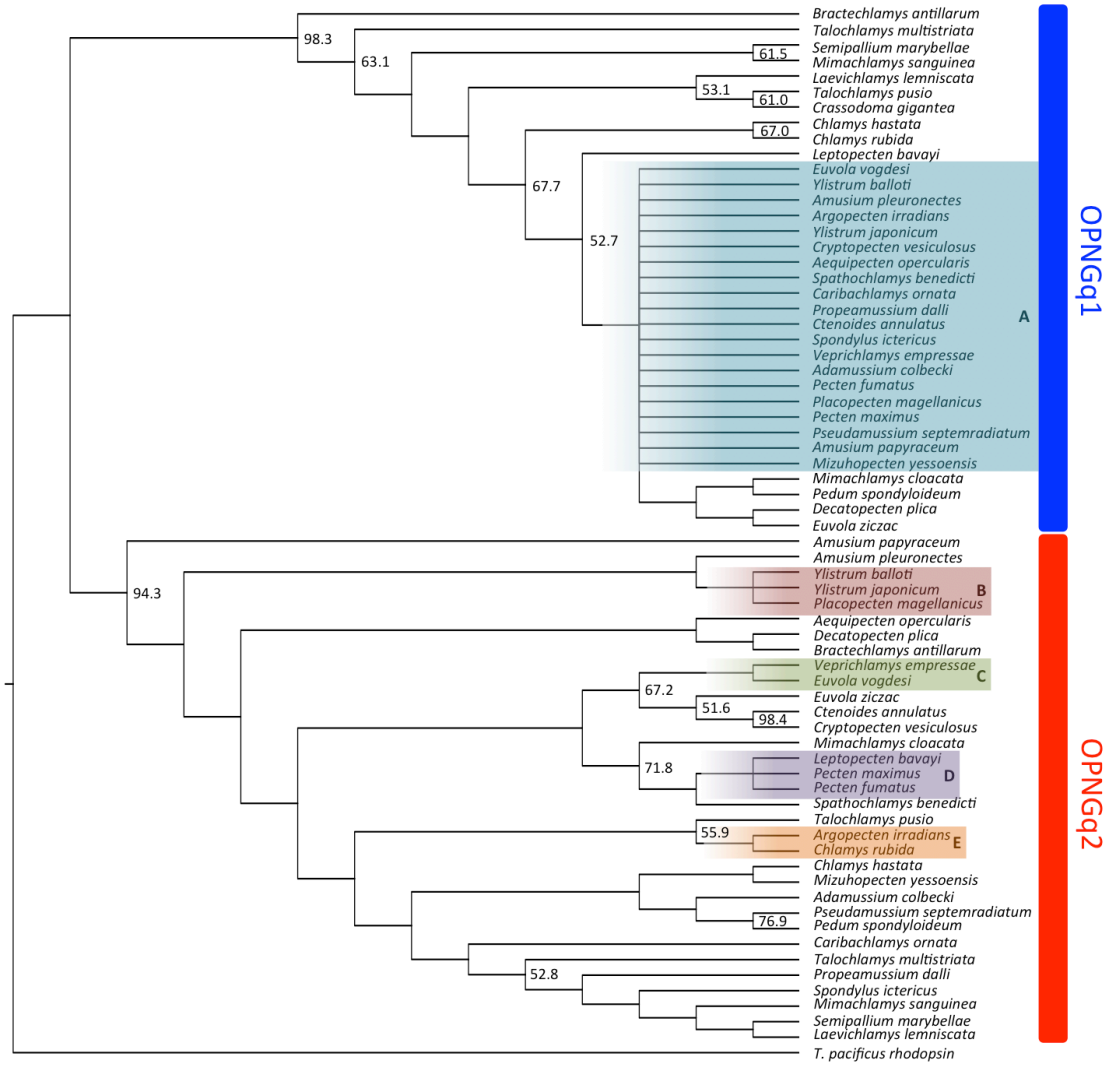


Figure 3-1 Maximum likelihood topology of G_q-opsin predominant amino acid sequence variants. The tree is based on aligned amino acid sequences. 34 taxa from the Superorder Pectenoida are represented by two clades, OPNGq1 (blue bar) and OPNGq2 (red bar). The outgroup is rhodopsin from a non-bivalve mollusc, *Todarodes pacificus*. Support values (>50%) of nodes were generated by 1000 bootstrap replicates in PhyML. Shaded taxa represent amino acid sequences that are identical to one another. Bolded letters A through E represent the predominant sequence variant that species' maintains.

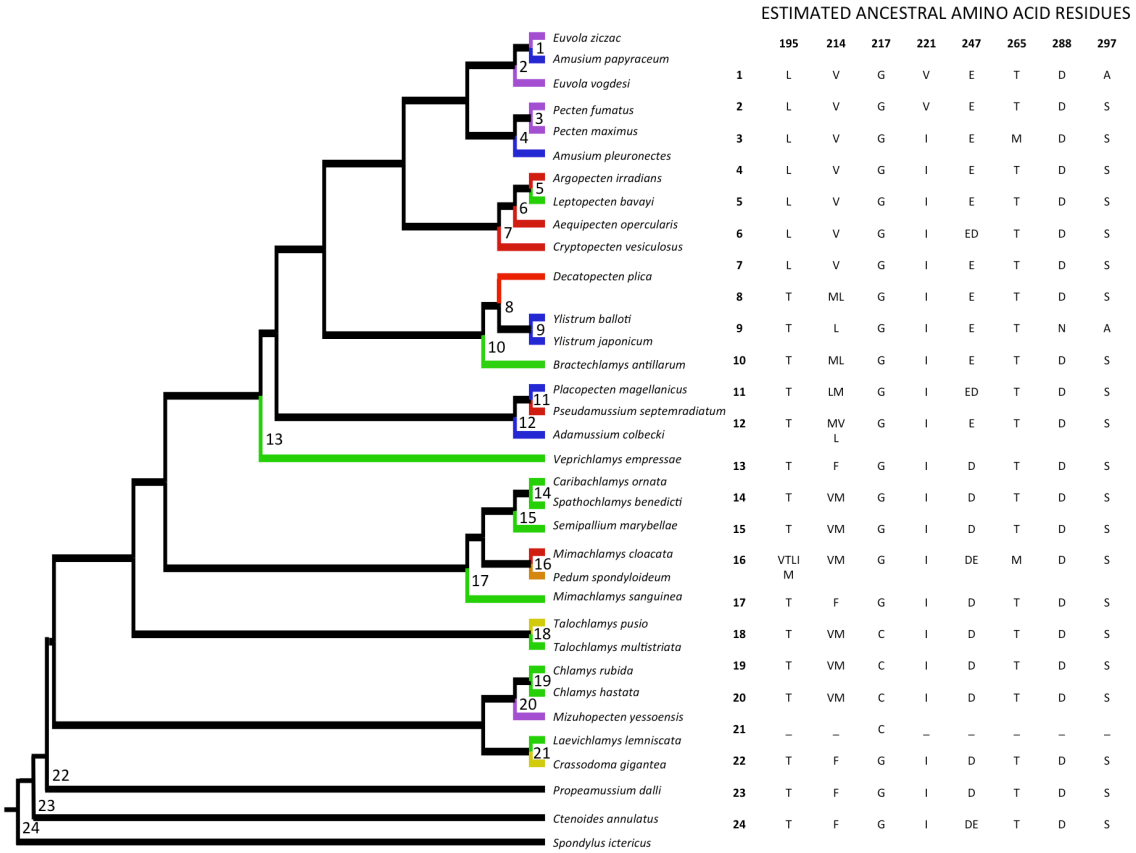


Figure 3-2 Estimated ancestral amino acid residues of scallop G_q-opsin proteins. Species topology is a trimmed majority rule consensus tree (Alejandrino *et al*., 2011) to include 34 species. Branch colors indicate behavior class (Alejandrino *et al.,* 2011). The estimated ancestral amino acid residues are listed at positions corresponding to Table 3-6; positions 195, 214, 221, 247, 265, 288, and 297 represent OPNGq2 and position 217 represents OPNGq1.

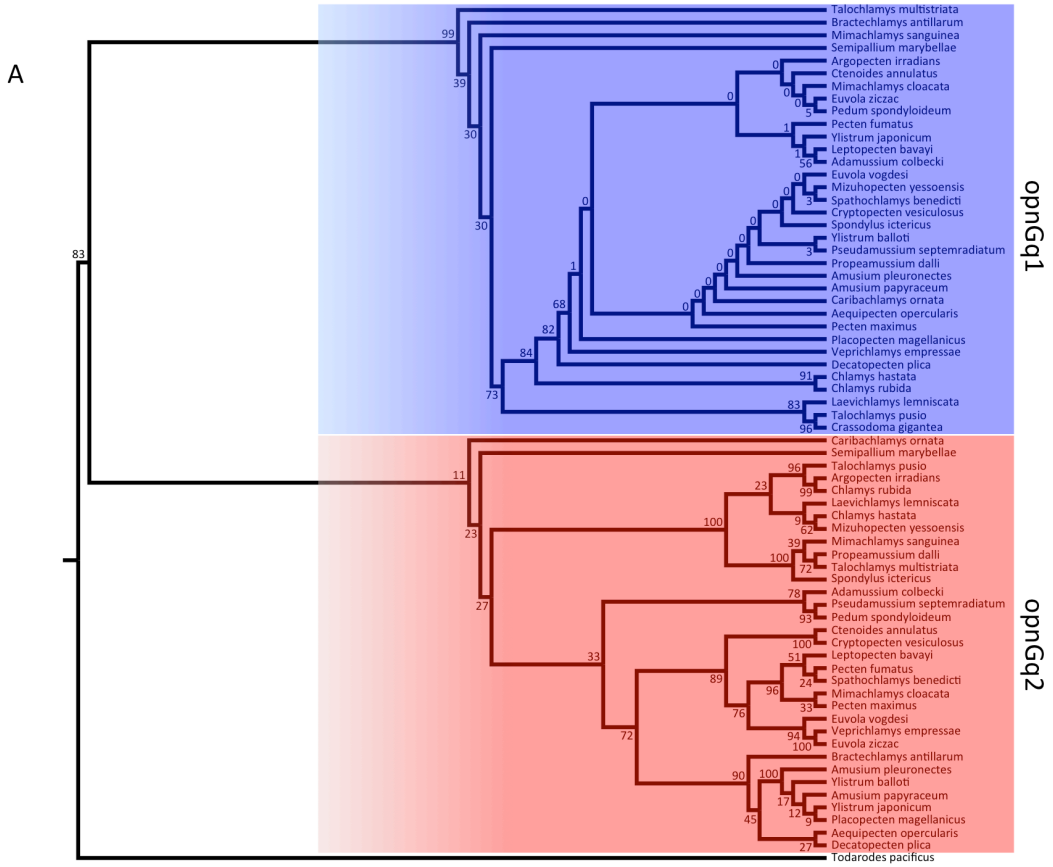


Figure 3-S1 CODEML tests. The maximum likelihood tree is based on aligned nucleotide sequences with *Todarodes pacificus* as the outgroup. Support values of nodes were generated by 1000 bootstrap replicates in PhyML. *opnGq1* sequences are colored red while opnGq2 sequences are colored blue. Panel A represents test 0 (T0), Panel B represents T1, Panel C represents T2, Panel D represents T3, Panel E represents T4, Panel F represents T5, and Panel G represents T6. A blue branch represents a branch with an omega value that is unequal to all other branches' omega values, which can also be called a release from selective constraint. A red branch represents a release from constraint that is a different from the blue release from selective constraint. Omega values from each test are given in Table 3-2.

Figure 3-S1 (continued)

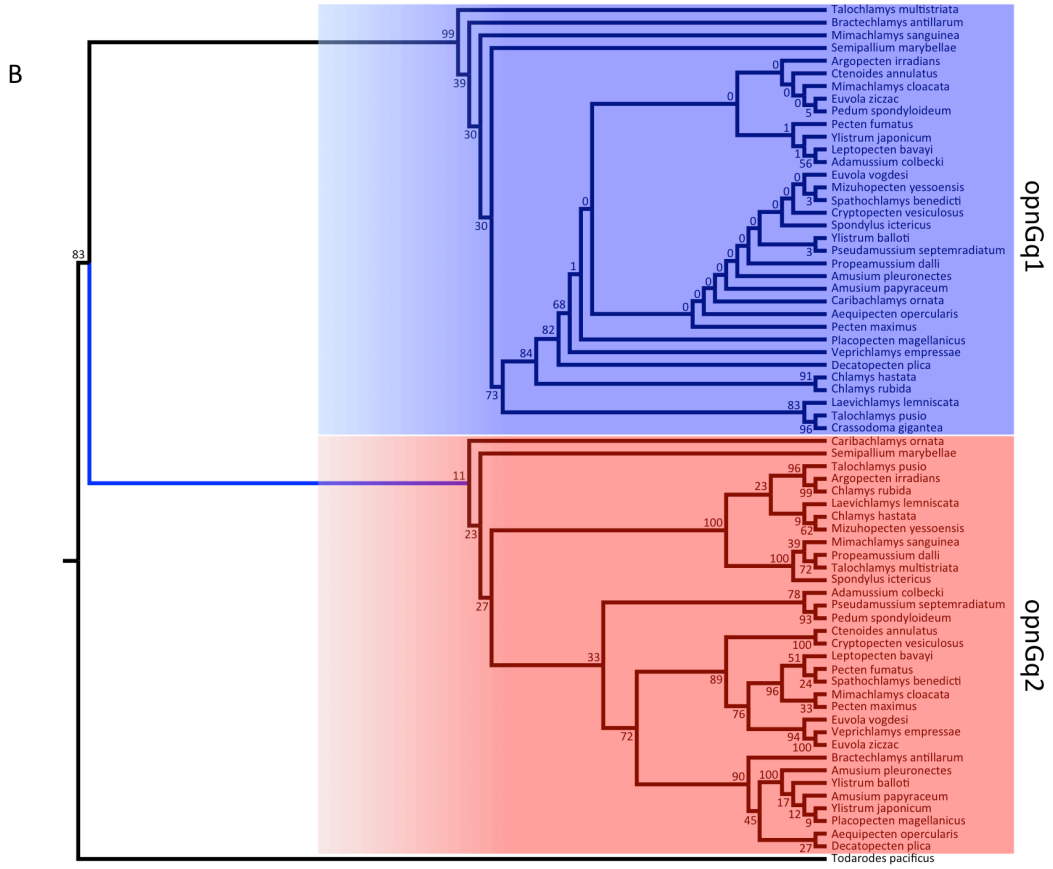


Figure 3-S1 (continued)

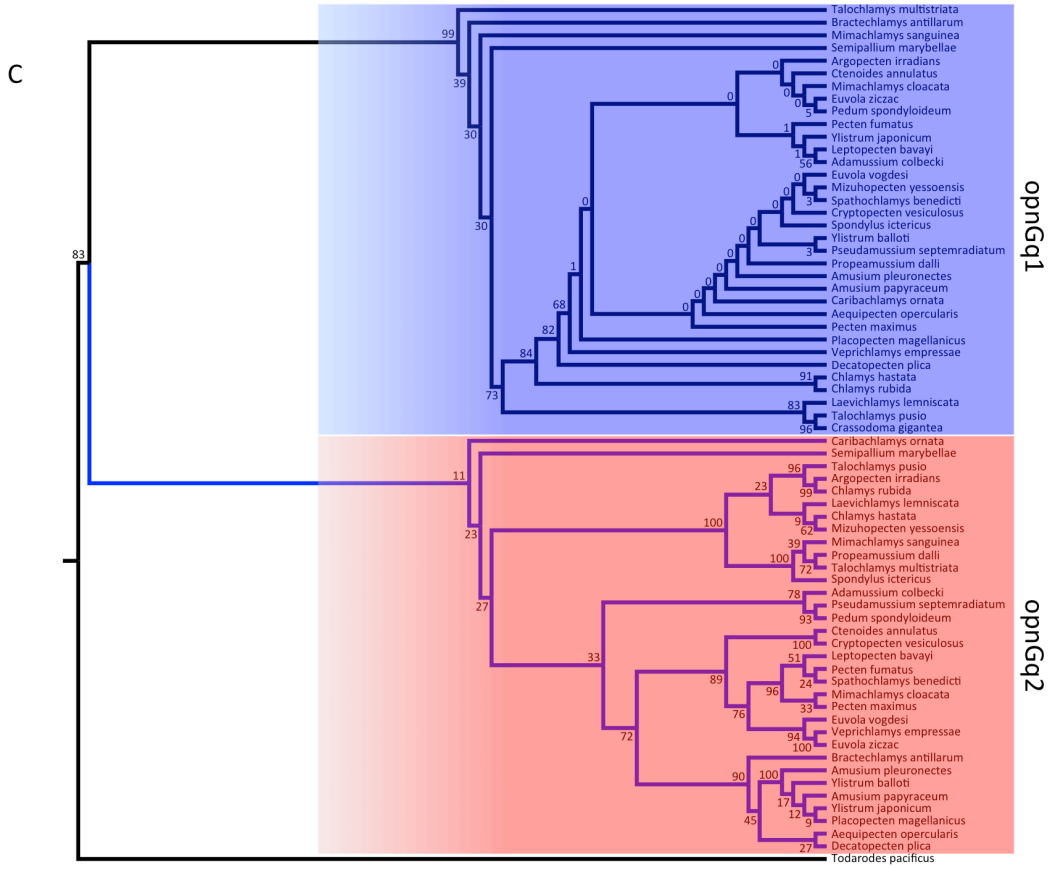


Figure 3-S1 (continued)

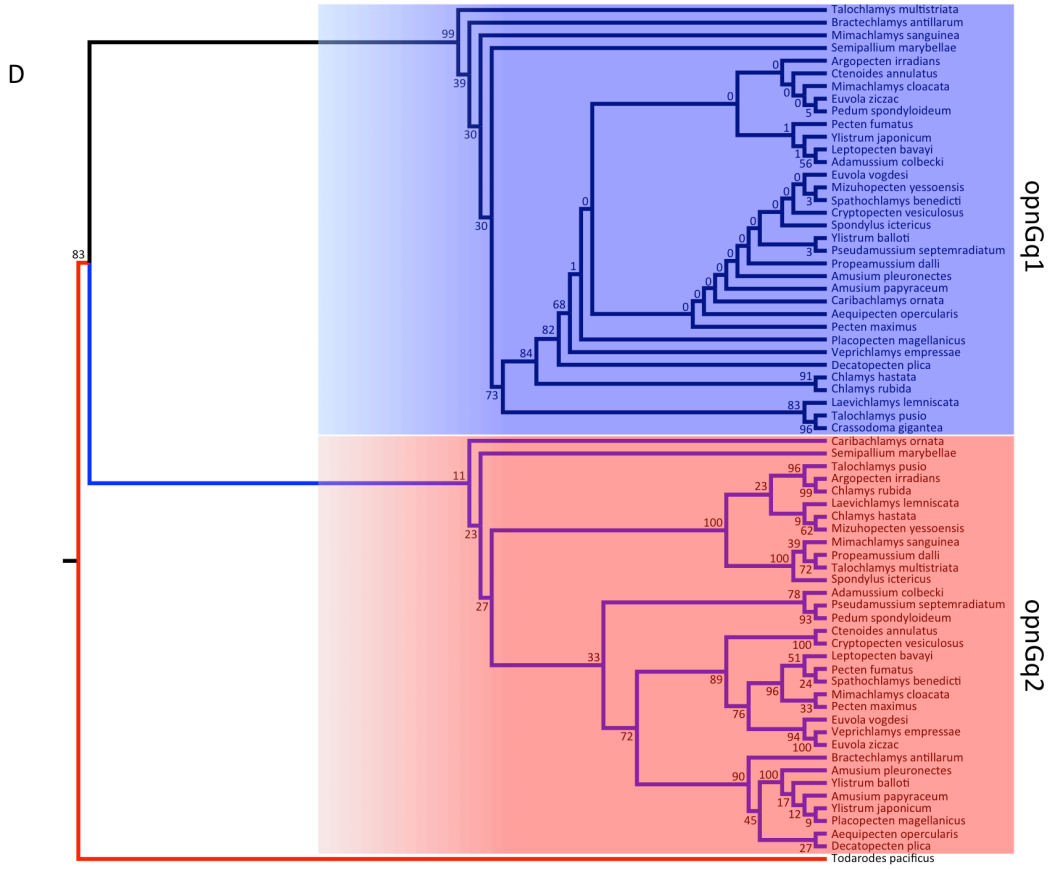


Figure 3-S1 (continued)

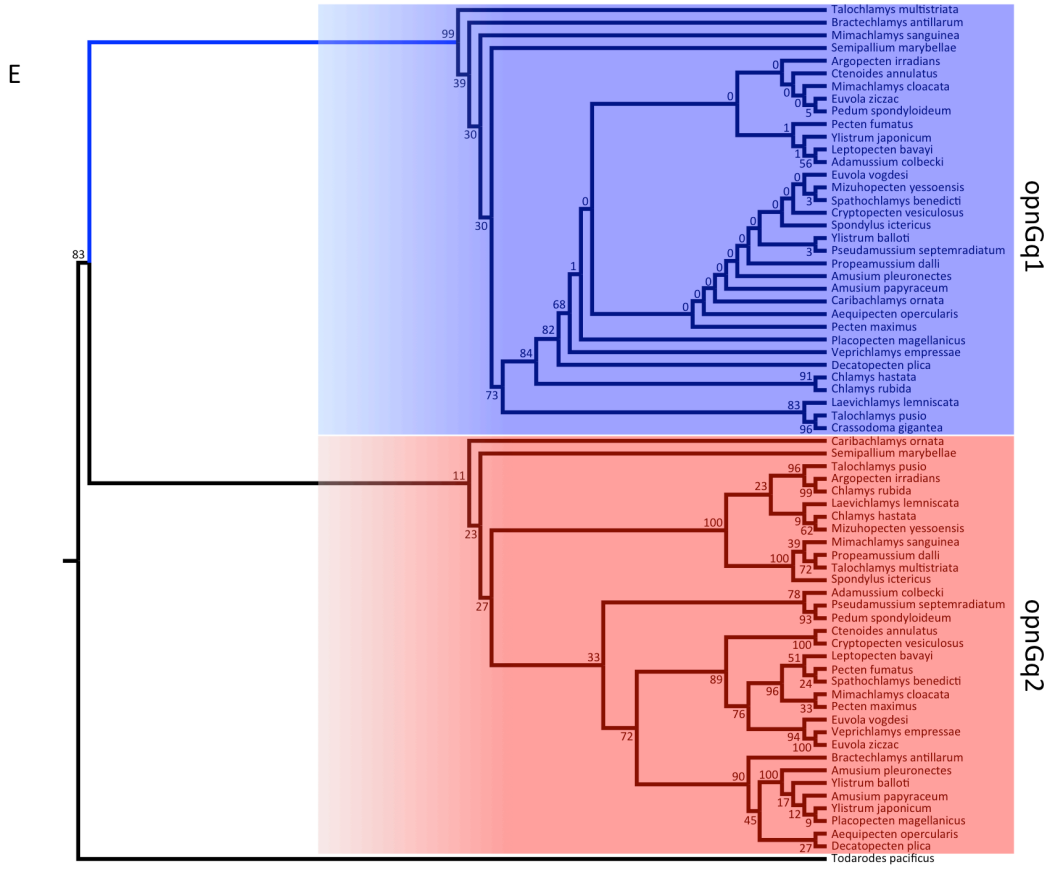


Figure 3-S1 (continued)

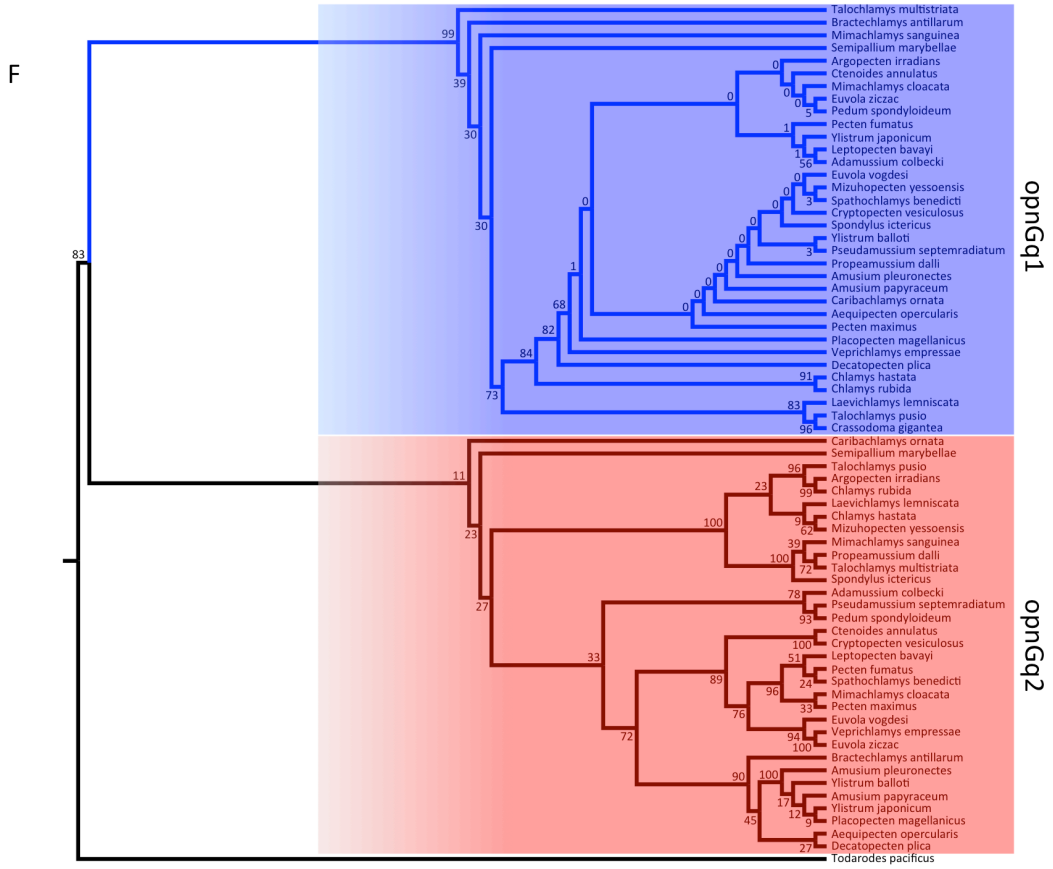
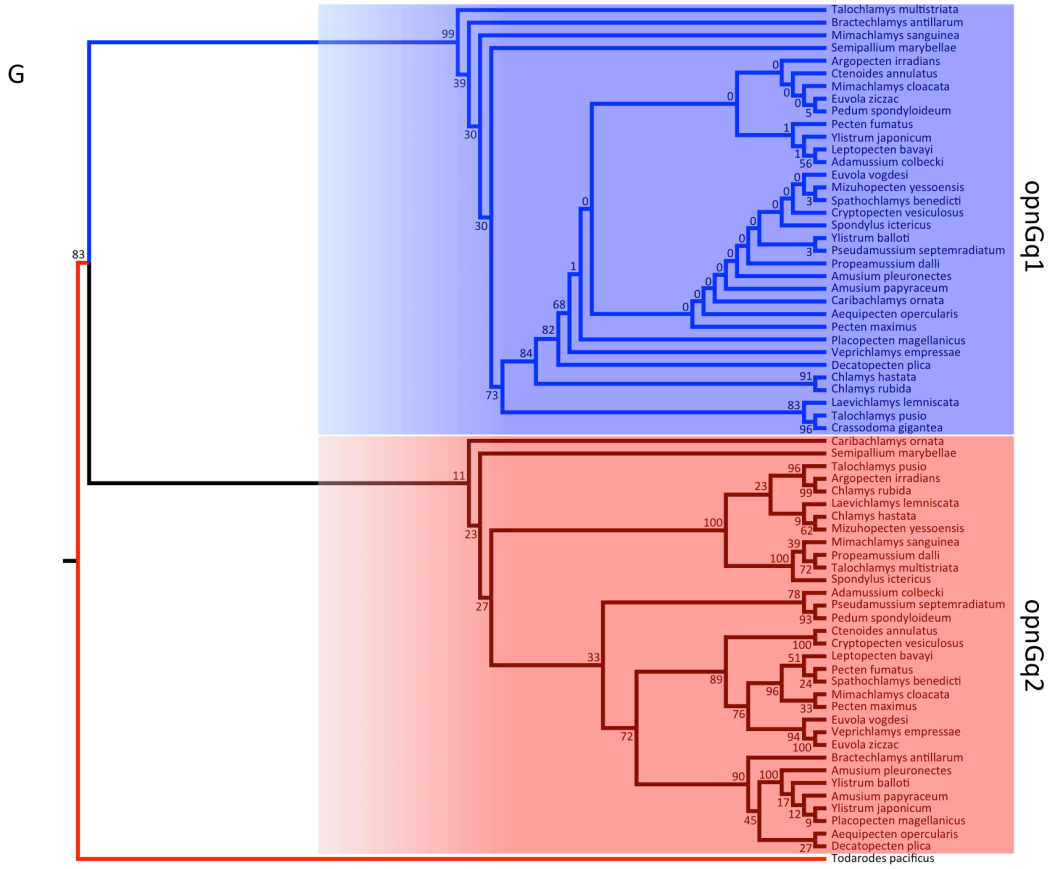


Figure 3-S1 (continued)



CHAPTER IV

THE DISADVANTAGE OF IGNORING ALLELIC VARIATION WITH AN EMPHASIS ON SCALLOP PHOTOSENSORY ALLELES

Anita Porath-Krause[1] and Jeanne M. Serb[1]

[1]*Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA*

Abstract

Allelic variation is a major contributor of genetic diversity and can cause drastic effects on phenotypes, yet these phenotypes may only represent a portion of the potential in a population. One type of allelic variation is cryptic genetic variation (CGV), is defined as standing genetic variation that will modify a phenotype that arises after environmental change or the introduction of novel alleles but does not contribute to the normal range of phenotypes observed in a population. This 'hidden' allelic variation can be advantageous to a population as it can facilitate an opportunity for adaptation to novel environmental conditions. In this paper, we highlight the important potential of cryptic genetic variation in the evolution of organisms; we reveal how alleles are often discarded in studies of molecular evolution; and, we emphasize the current role of allelic variation described in sensory systems belonging to GPCRs. To demonstrate the breadth of possible allelic variation within an individual, we report on the allelic variation in G_q-opsin sequences, a component of photosensory systems, sampled across 34 bivalve species.

Introduction

Genetic diversity has profound effects on the survival and evolution of organisms. It is critical for the ability to cope with changing environments, adapt to local environments, and increase resistance to extinction in natural populations (Webster *et al.*, 1972; Yang & Patton, 1981; Barrett & Schluter, 2008). Alternative forms of a gene at a locus, known as allelic variation (Graur & Li, 2000; X *et al.*, 2002), is a major contributor of genetic diversity (Graur & Li, 2000; X *et al.*, 2002; Costello *et al.*, 2003; Mock *et al.*, 2013) in natural populations of diploid organisms. Classic examples of allelic variation include easily distinguishable phenotypic differences, such as body color (Pointer & Mundy, 2008; Rosenblum *et al.*, 2010; Hoekstra *et al.*, 2006; Grant *et al.*, 1996); but not all phenotypic differences are easily detected (West-Eberhard, 1983). Allelic variation has been shown to affect organisms phenology (Johanson *et al.*, 2000), physiology (Day *et al.*, 1974), or continuous traits (Todesco *et al.*, 2010); and can even affect phenology at the microscopic level. For example, human individuals carrying a single allele for the sickle cell phenotype experience reduced severity of malaria and increased likelihood of survival when infected with *Plasmodium falciparum* due to a heterozygous advantage (Lell *et al.*, 1999). The sickle cell phenotype is not easily observed by the naked eye or simple tools, but can be quantified in terms of survival for the individual.

Current use of allelic variation

Cryptic genetic variation plays an important role in evolution

One form of allelic variation, known as cryptic genetic variation (Muller, 1949), may have important genetic potential for the survival and evolution of organisms. In the

context of this paper, cryptic genetic variation is defined as genetic variation that is maintained in the form of allelic variation through stabilizing selection (reviewed in Paaby & Rockman, 2014), but the allele is not expressed until the population is under environmental duress with atypical or extreme environmental conditions (Hayden *et al.*, 2011; Paaby & Rockman, 2014). The 'hidden' allelic variation can be advantageous to a population as it can facilitate an opportunity for adaptation to novel environmental conditions. Cryptic genetic variation has been shown to contribute to the evolution of complex traits (e.g., eye size in Mexican cave compared to surface dwelling tetra fish Rohner *et al.*, 2013), body size (e.g., freshwater diatoms grown in low salinity conditions Beszteri *et al.*, 2005), and sex determination (e.g., cyclic parthenogens of cladocerans subjected to additional growth hormones Kim *et al.*, 2006). Clearly playing an important role in the response to new environments, cryptic genetic variation is proposed to be so pivotal for adaptation that is was vital during historical evolutionary events like the transition of uni- to multicellularity (Schlichting, 2008).

Unexplained allelic variation is frequently discarded in molecular studies

Allelic variation can cause drastic effects on phenotypes, yet is frequently discarded or simplified in molecular studies. Whether a lack of statistical robustness, insufficient computational capabilities, or a misunderstanding of the raw data, it is unclear why many studies exclude allelic variation from their data. Below, we describe three examples when allelic variation was excluded from molecular evolutionary studies; yet, the presence of multiple alleles was explicitly reported, which can be more useful and informative to the reader than not reporting the findings. Each example involves G-

protein coupled receptors (GPCR), which are a large family of membrane receptors that bind to a ligand and initiate signal transduction within the cell. The amino acid sequence dictates the folding, and thus shape, of the GPCR's tertiary structure; therefore nonsynonymous changes in the nucleotide sequence have the potential to directly alter the function of the receptor (Yokoyama & Radlwimmer, 1999; Parry *et al.*, 2004; Kim & Drayna, 2004). GPCRs have a one-to-one genotype to phenotype relationship meaning changes at the level of the gene potentially have an immediate, direct affect on the corresponding phenotype, which makes further tests of correlation and adaptation relatively straightforward. One study looked at the molecular evolution of duplicated color genes (LWS-1 and SWS2) in cyprinid fish and found multiple sequences with polymorphic base pairs were isolated from the same fish specimen using gene-specific primers (Li *et al.*, 2009). Of the sixteen cyprinid species tested in the study, thirteen had ten or more different SWS2 sequences captured in cloning using PCR-based methods (Li *et al.*, 2009). The authors chose to reduce the number of allelic variants included in further analyses to one or two representative sequences that contained the least number of autapomorphies to make interspecific comparisons. They recognized that the different sequences could be representative of different alleles; however, the authors also speculated that the sequence differences could be the result of errors during PCR amplification. According to TaKaRa (Takara Bio Company © 2016), the quality of *Ex Taq* ® polymerase used in the study has a mutation rate of approximately 4.5 times lower than standard Taq® polymerase (standard Taq® polymerase = 68.4% of PCR products of 1kb length will have one error), suggesting the polymorphisms identified between SWS2 sequences are most likely real and not the result of base misincorporations during PCR

amplification. A similar pattern was seen in the genome of the pea aphid (*Acyrthosiphon pisum*), which was used to look for host plant specialization based on chemosensory processes when compared to other arthropod chemoreceptors (Smadja *et al.*, 2009). A total of 77 genes from the gustatory receptor family and 79 genes from the odorant receptor family were identified in the pea aphid genome. An additional seven variants from the gustatory receptor family and one variant from the odorant receptor family were identified in the pea aphid. While these variants were identified as alleles belonging to the chemoreceptor families, the eight allelic variants were excluded from phylogenetic analyses and tests of selection. The authors did not explicitly state the reason for removing allelic variants from further analyses but if the alleles were included, they may have revealed even greater genetic expansion of these chemoreception genes in the genome. In the third example, a study using humans, bonobos, and chimpanzees searched for sequence variability among bitter taste receptors, T2R, to explain the source of personalized taste perception (Parry *et al.*, 2004). The T2R genes were isolated and amplified through PCR using gene-specific primers, cloned, and then at least three clones per T2R gene isolation were sequenced. The authors explicitly stated that if there were differences among clones, additional sequencing was executed until there was a consensus sequence. The consensus sequence was used in further comparisons and the non-consensus sequences were excluded from further analyses. Again, the authors did not explicitly state their reason for removing allelic variants, but if the additional variants remained in the study, the degree of sequence variability between T2R genes may have been much greater, suggesting that personalized taste variation is even more diverse than what is recorded in the study.

Opsins are a robust model to describe importance of allelic variation

The current approaches used to test allelic variation in GPCRs (described below) may not be sufficient for all systems to address the breadth of molecular evolutionary questions. Using one or two alleles may be adequate to study most gustatory (i.e. taste) sensory systems in vertebrates which commonly utilize psychophysical tests (Bufe *et al.*, 2005; Faurion *et al.*, 1980; Duffy & Bartoshuk, 2000; and reviewed in Reed *et al.*, 2006) and expression studies (Matsunami *et al.*, 2000; Miyoshi *et al.*, 2001) to determine individual taste preference; but, when expanding analyses to include tests such as rates of evolution, estimating forces of selection, and reconstructing phylogenetic relationships, the inclusion of all allelic variation may be critically important to the analysis, e.g., increasing or decreasing dN/dS values or changing phylogenetic relationships. One particular sensory system of interest utilizes the Class A GPCRs, which comprise the largest group of GPCRs and include the light-sensitive protein, opsin. Housed in the photoreceptor cell, opsins are seven-transmembrane proteins that act as a visual pigment when covalently bound to a light-sensitive chromophore. Opsins have long been the focus of many papers testing molecular evolutionary patterns, rates of evolution, and estimating forces of selection (Yokoyama, 2002; Dann *et al.*, 2004; Porter *et al.*, 2006; Frentiu *et al.*, 2007; Carleton *et al.*, 2010); for example, interspecific comparisons of opsin absorbance can reveal patterns of adaptation and convergence of wavelength absorption values (Yokoyama *et al.*, 2008). Opsin wavelength absorption values are quantified by the wavelength at which the absorbance of light is the greatest (λ_{max}). The λ_{max} value is a phenotype which can be quantified directly with tools that measure a

discrete value (Terakita *et al*., 2008; Townson *et al*., 1998; Feiler *et al*., 1992; Oprian *et al.*, 1987; Asenjo *et al.*, 1994; Wilkie *et al.*, 2000; Sakmar *et al.*, 2002). The individual phenotypic variation (λ_{max}) can then be directly compared to other λ_{max} values (Arrese *et al.*, 2002; O'Connor *et al.*, 2010; Speiser *et al.*, 2011) and have proven to contribute greatly to addressing questions of molecular evolution (Yokoyama, 2002; Takahashi & Ebrey, 2003; Yokoyama *et al.*, 2008). If a non-synonymous allelic variant causes a change in the tertiary structure, the resulting phenotype may be detected by these sensitive techniques; therefore, it would appear exceptionally informative to include all allelic variants in tests of molecular evolution.

Scallops as a model to call attention to allelic variation

In our lab, we are developing a new model system, the scallop, to explore the role of allelic variation in molecular evolution of opsin. The eyes of scallops have been of particular interest because of the double retina structure of the eye that contains at least two classes of opsin, G_q-opsin and G_o-opsin (Kojima, 1997). In two scallop species, the wavelength absorption was measured in both the G_q-opsin and G_o-opsin and confirmed that G_q-opsin and G_o-opsin produce a different λ_{max} value in each species (Speiser *et al.*, 2011). In addition to their unique eye morphology, the scallop G_q-opsin has been relatively well-characterized (Speiser *et al.*, 2011; Kojima, 1997; Nasi & Gomez, 2009; Terakita *et al*., 1993; Speiser & Johnsen, 2008). For example, electrophysiological investigations show that G_q-opsins utilize a photo transduction cascade that results in a depolarization (decrease in membrane potential) of the photoreceptor cell (Nasi & Gomez, 2009), and studies using immunohistochemistry show that the $G_{q\alpha}$ subunit

colocalizes with the G_q-opsin in the depolarizing cells of the scallop proximal retina (Kojima, 1997). A recent study highlights the molecular relationship among G_q-opsin gene paralogs revealed 30 scallop species and three species sampled from closely allied marine bivalve families also representing the Pectinoida suborder, Propeamussiidae, Limidae, and Spondylidae, contained two putatively functional G_q-opsin paralogs in each species (Porath-Krause & Serb, in prep). The two G_q-opsin paralogs, *opnGq1 and opnGq2,* were shown to have different evolutionary trajectories across Pectinidae with *opnGq2* under stronger purifying selection than *opnGq1*. Porath-Krause & Serb (in prep) identified three or more *opnGq1* nucleotide sequences in 21 scallop species and three or more *opnGq2* nucleotides sequences in 17 species when isolating G_q-opsin paralogs (see Methods in Porath-Krause & Serb, in prep); however, many allelic variants were excluded from further analyses to compare interspecific, rather than intraspecific, variation. In this study, we leverage those allelic variants to demonstrate the importance of including all allelic variation in tests of molecular evolution.

Methods

To capture the magnitude of potential allelic variation present in each scallop species, we isolated two G_q-protein coupled opsin genes (herein *opnGq1* and *opnGq2* for the gene or the coding region, and OPNGq1 and OPNGq2 for the protein) from 31 scallops sampled across the scallop phylogeny. In addition, we chose a single species from each of the three closely allied bivalve families (Figure 1 in Alejandrino *et al.*, 2011), Propeamussidae, Limidae, and Spondylidae, which are also known to possess both *opnGq1* and *opnGq2* (Serb *et al.*, 2013). To isolate the nucleotide sequences for the two

G_q-opsin genes, ethanol-preserved tissues from the Pectinoida were obtained from museum collections or provided by colleagues (see Table 1 in Puslednik & Serb, 2008; Alejandrino *et al.*, 2011). Total genomic DNA was extracted from these samples following the methods described by Puslednik and Serb (Puslednik & Serb, 2008). Polymerase chain reaction (PCR) was performed using published gene-specific primers and methods described in Serb *et al*., 2013 (Table 2) to amplify a 492 base pair region for *opnGq1* and a 489 base pair region for *opnGq2*. This region represents approximately 35% of the entire coding region and contains the lysine motif that binds the chromophore in the seventh transmembrane (TM7). PCR products were cloned into TOPO TA vector that was used to transform *E. coli* following the manufacturer's instructions (TOPO TA Cloning Kit with pCR2.1-TOPO). Colonies were blue-white screened and at least ten white colonies containing recombinant DNA were selected. We amplified the recombinant DNA from each colony using M13 forward and M13 reverse primers (specified by the TOPO TA Cloning Kit) according to manufacturer's instruction. We confirmed that the plasmid contained an insert of the target size using a 1% agarose gel electrophoresis. Clones with the correct insert size were sequenced with an ABI 3730 Capillary Electrophoresis Genetic Analyzer at the Iowa State University DNA Sequencing Facility. The resulting DNA sequences were translated and manually checked for stop codons and the presence of the xAKxS motif in TM7 indicating chromophore-binding capability of the translated protein. DNA sequences that did not meet these requirements were considered nonfunctional pseudogenes and were removed from further analyses. Nucleotide sequences were blasted in NCBI to confirm gene identity as either *opnGq1* or *opnGq2* (described as Gq-opsinB and Gq-opsinA,

respectively, Serb *et al.*, 2013). Any identical nucleotide sequences that were collected from the same species were not included in further analyses. All newly generated sequences were deposited in NCBI Genbank database (accession numbers to be entered after submission) (Table 4-1). To expose evidence of allelic variation among nucleotide sequences producing nonsynonymous amino acid sequences, the translated sequences were aligned by MAFFT v.7.017 (Katoh *et al.*, 2002) in Geneious Pro v. 5.6.7 (Drummond *et al.*, 2011) using default parameters. Unique and identical amino acid sequences within and between species were recorded. In addition, we trimmed a comprehensive scallop species tree (Alejandrino *et al.*, 2011) to include only those 34 taxa representative of the *OPNGq1* and *OPNGq2* sequences. For each species, all possible amino acid sequences are presented in the form of pie charts, with each unique amino acid sequence having a different color.

To identify the mutations causing allelic variation within a species, each species that contained greater than one representative *opnGq1* or *opnGq2* nucleotide sequence were aligned by MAFFT v.7.017 (Katoh *et al.*, 2002) in Geneious Pro v. 5.6.7 (Drummond *et al.*, 2011) using default parameters. The total number of single nucleotide base substitutions within the aligned sequences was recorded. A pairwise percent identity between each combination of two aligned nucleotide sequences was recorded, as well.

To demonstrate the stochastic effect of excluding alleles from tests of molecular evolution, we randomly selected one nucleotide sequence from each species. *opnGq1* and *opnGq2* were randomized and tested separately from one another. Omega was then calculated using the PARRIS method (Scheffler *et al.*, 2006) in DataMonkey (Pond &

Frost, 2005; Delport *et al.*, 2010). Quantifying omega, the ratio of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site (ω = dN/dS rate ratio) is a common way to test the extent of selection on sequence evolution. The PARRIS method tests for evidence of selection in an alignment-wide framework. PARRIS applies a partitioning approach to sequences to infer separate branch lengths and topologies to each partition; so, within each partition, any site-to-site variation in the synonymous rate is accounted for. The dN/dS rate ratio (ω) was calculated three times for *opnGq1* and *opnGq2*, each from nucleotide sequences that were randomized from the nucleotide sequences available in each species.

Results

Allelic variation was prevalent among the 34 species sampled for both G_q-opsin genes. From the 34 species sampled, between one and six alleles of *opnGq1* and *opnGq2* were identified in each species, with the exception of *Crassadoma gigantea* (Table 4-1), from which we were not able to clone any *opnGq2* sequences. For both *opnGq1* and *opnGq2*, the most frequent number of different alleles per species was four. One, two, and three alleles were more common than five or six alleles per species. Only five species had a single allele that represented *opnGq1* and seven species had a single allele representing *opnGq2*. The pattern of containing a single allele was not shared between any *opnGq1* and *opnGq2* that were sampled from the same species.

Within the same G_q-opsin paralog, we found a wide range of genetic variation when comparing two or more sequences within a species. There were 105 different *opnGq1* alleles and 99 *opnGq1* alleles found across all 34 species sampled (Table 4-1).

When comparing the number of single nucleotide polymorphisms between two or more *opnGq1* alleles from the same species, 32 of the 34 species had approximately 98 percent or greater pairwise identity (Table 4-2). Twenty-seven species had between one and 15 single nucleotide polymorphisms (SNPs) between two or more *opnGq1* allele sequences. Two species, *Leptopecten bavayi*, and *Caribachlamys ornata*, had 61 and 70 SNPs, respectively. For the 492 base pair long sequence, this is approximately a 12 to 14 percent difference between sequences. *opnGq2* had much more allelic variation within each species with only 14 species that contained two or more nucleotide sequences at approximately 98 percent or greater pairwise identity. These species had between one and nine SNPs among nucleotide sequences. More than one-third of the Pectinoida species had a large number of SNPs, between 16 and 116, among *opnGq2* nucleotide sequence with a pairwise identity of less than 98 percent. For the 489 base pair long sequence, this ranges from three to approximately 24 percent difference between sequences.

Based on which alleles were chosen, the omega values for *opnGq1* and *opnGq2* changed. The interspecific dN/dS ratio was calculated from a randomized sampling of all available alleles three times for both *opnGq1* and *opnGq2* and showed that dN/dS ratios change based on which alleles are included in the analysis. In *opnGq1*, the dN/dS ratio was calculated to be as low as 0.06 in the second calculation, and as high as 0.10 in the first run (Table 4-3). While these two values are both considered evidence of purifying selection, they are different. *opnGq2* did not have as large of a difference between randomized runs with the lowest dN/dS ratio of 0.06 and highest of 0.07.

Not only is there a great abundance of *opnGq1* and *opnGq2* nucleotide sequence diversity within many of the species sampled, the majority of those nucleotide sequences frequently translated to different amino acid sequences within a species (Figure 4-1). Eighteen species containing more than one *opnGq1* and 15 species containing more than one *opnGq2* nucleotide sequences have nucleotide sequences that translate to an equal number of unique amino acid sequences (Table 4-1). One species, *Talochlamys multistriata,* has as many as six different *opnGq1* nucleotide sequences that translate to six OPNGq2 amino acid sequences.

Discussion

Our study exemplifies the most comprehensive representation of sequence diversity found across Pectinidae. We included all alleles identified for *opnGq1* and *opnGq2* and highlighted the breadth of nucleotide variation and corresponding nonsynonymous amino acid sequence variation found within a species. SNPs have the potential to cause major functional differences so evidence of divergence, convergence, and phenotypic evolution at the level of the protein may be lost if alleles are excluded from analyses of molecular evolution. Here, we stressed the unintentional bias that could result from ignoring allelic variation found within an individual.

The high number of nucleotide sequences in many species was unexpected. We only anticipated a maximum of two alleles to be present for each gene in each species because unless the animals are induced through aquaculture practices to become triploid (Tabarini, 1984), most extant scallops that have been karyotyped are diploid organisms (reviewed in Odierna *et al.*, 2006) and are expected to have two alleles per gene.

However, others have hypothesized that the ancestral karyotype for the family Pectinidae is tetraploid, resulting from whole genome duplication (Wang & Guo 2004). While we can only speculate, a whole genome duplication event followed by chromosomal reduction due to fusion and arm loss of telocentric chromosomes may have allowed multiple copies of the same gene to be maintained which could explain the number of divergent nucleotide sequences. Many of our allelic variants were caused by a few SNPs, which is not unusual to detect within a species (Flordellis *et al.*, 2004; Hoekstra *et al.*, 2006). In the guppy, *Poecilia reticulata*, for example, 15 alleles were found to encode seven long-wave sensitive (LWS) opsin protein variants. The nonsynonymous alleles were the result of up to six nonsynonymous SNPs (Hoffmann *et al.*, 2007). However, the LWS clones were sampled from guppies collected from nine different strains (or different localities) and the differences could be due to either individual or geographic differences. In our study, however, we also identified multiple different nucleotide sequences with several SNPs, which were sampled from a single individual. The degree of allelic variation within a single Pectinoida individual appears to be unprecedented.

Caution should be taken when excluding alleles from analyses if dN/dS ratios are used in tests of molecular evolution that highlight selective regime or strength of selection (e.g., Dann *et al.*, 2004; Spady, 2005; Li *et al.*, 2016). For example, in a test of molecular evolution of two paralogous circadian genes in eukaryotes, the relative strength of selection was compared between the two genes *timeless* and *timeout*. *timeless* was found to be under stronger selective pressure ($\omega = 0.02182$) than *timeout* ($\omega = 0.13955$) (Li *et al.*, 2016). If a similar comparison was made between *opnGq1* and *opnGq2* using a representation of alleles with sequences that were excluded, *opsGq1* could be considered

under stronger selection (ω of run 2 = 0.06) than *opsGq2* (ω of run 3 = 0.07). Alternatively, if the comparison made between *opnGq1* and *opnGq2* used a different representation of alleles, the outcome would be the opposite. The strength of selection would be stronger in *opnGq2* (ω of run 2 = 0.06) than in *opnGq1* (ω of run 1 = 0.10), revealing how different representations of allele sequences in calculations of dn/ds can bias interpretations of relative strengths of selection.

Evidence suggests we uncovered hidden variation within a species. We identified nucleotide sequences that translate to a diversity of OPNGq1 and OPNGq2 alleles. This pattern of high amino acid variation may not be an anomaly to this study; from the Li et al. cyprinid fish study described earlier, *LWS-1* formed two well-supported monophyletic lineages. After translation, the LWS-1 amino acid sequences were aligned and showed amino acid differences among sites within the same species. It is possible that if all nucleotide sequences from *LWS-1* were included in the final analyses, there may be an abundance of amino acid sequence variation within a species. We posit that the alleles we identified in Pectinoida may remain silent in the form of cryptic variation, only to be expressed as phenotypic variation when the scallop population is under atypical environmental changes. Scallops live in relatively ephemeral and heterogeneous environmental light conditions, with changing water clarity, depth, and light intensity (Lythgoe, 1968; Jerlov, 1976). Our suspected cryptic variation may facilitate photosensory adaptation to rapidly shifting light environments of the ocean which are led by contemporary climate induced changes (Behrenfeld *et al.*, 2006).

Most examples discussed here highlight how allelic variation may affect phenotype or function; nonetheless, some researchers use theory to posit that allelic

variation could have no effect on the phenotype because phenotypically neutral alleles that are not affected by natural selection can accumulate in a population (Kimura & Ohta, 1973; Ewens & Gillespie, 1974). While there are rarely empirical examples (e.g., Shi *et al.*, 2009), many papers discuss the theory, supported by models, behind neutral alleles in allelic variation (reviewed in Lynch & Hill, 1986). The allelic variation observed in G_q-opsin across scallops may be due to a random accumulation of mutations. Above, we discussed how the nucleotide sequences that produced a diversity of OPNGq1 and OPNGq2 alleles are potentially examples of cryptic variation that produce undetected phenotypes; however, we do not directly measure the λ_{max} values of each allele so we are unsure whether the nonsynonymous changes affect the phenotype. The observed amino acid differences could be functionally neutral. Alternatively, because multiple species contain numerous nonsynonymous sequences, a portion of the sequences could be neutral while a portion could change the λ_{max} value. Whether the nucleotide sequence variation is neutral or non-neutral, it may be critical to include and explore the role of all alleles in tests of molecular evolution. We demonstrated that by eliminating alleles or randomly selecting alleles in calculations of the ratio of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site could skew resulting dN/dS ratios and possibly lead to further misinformed tests of molecular evolution.

Why is there so much *opnGq* allelic variation within a species? We expect the presence of *opnGq1* and *opnGq2* in most scallop species, but further abundance of nucleotide sequences found within each gene suggests at least one additional round of duplication events. We cannot ignore that we did not sequence the complete coding

region for each allele so we cannot confirm the absence of stop codons upstream or downstream of the coding region we captured. Additionally, we used standard Taq® polymerase so errors due to point mutations may introduce false allelic variants. While Taq® errors are possible at a rate of one error per every one kilobase sequence length (expect two per ~500bp region), we argue that this could not explain the vast number of species with greater than 50 SNPs between alleles making the molecular relationships in this scallop photosensory system extremely complicated and ripe for further studies.

References

Alejandrino, A., Puslednik, L. & Serb, J.M. 2011. Convergent and parallel evolution in life habit of the scallops (Bivalvia: Pectinidae). *BMC Evol. Biol.* **11**: 164.

Arrese, C.A., Hart, N.S., Thomas, N., Beazley, L.D. & Shand, J. 2002. Trichromacy in Australian Marsupials. *Curr. Biol.* **12**: 657–660.

Asenjo, A.B., Rim, J. & Oprian, D.D. 1994. Molecular determinants of human red/green color discrimination. *Neuron* **12**: 1131-1138.

Barrett, R.D.H. & Schluter, D. 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol. (Amst.)* **23**: 38–44.

Behrenfeld, M.J., O'Malley, R.T., Siegel, D.A., McClain, C.R., Sarmiento, J.L., Feldman, G.C., *et al.* 2006. Climate-driven trends in contemporary ocean productivity. *Nature* **444**: 752–755.

Beszteri, B., Ács, É. & Medlin, L.K. 2005. Ribosomal DNA Sequence Variation among Sympatric Strains of the Cyclotella meneghiniana Complex (Bacillariophyceae) Reveals Cryptic Diversity. *Protist* **156**: 317–333.

Bufe, B., Breslin, P.A.S., Kuhn, C., Reed, D.R., Tharp, C.D., Slack, J.P., *et al.* 2005. The Molecular Basis of Individual Differences in Phenylthiocarbamide and Propylthiouracil Bitterness Perception. *Curr. Biol.* **15**: 322–327.

Carleton, K.L., Hofmann, C.M., Klisz, C., Patel, Z., Chircus, L.M., Simenauer, L.H., *et al.* 2010. Genetic basis of differential opsin gene expression in cichlid fishes. *J. Evol. Biol.* **23**: 840–853.

Costello, A.B., Down, T.E., Pollard, S.M., Pacas, C.J. & Taylor, E.B. 2003. The influence of history and contemporary stream hydrology on the evolution of genetic diversity within species: an examination of microsatellite DNA variation in bull trout, Salvelinus confluentus (Pisces: Salmonidae). *Evolution* **57**: 328–344.

Dann, S.G., Allison, W.T., Levin, D.B., Taylor, J.S. & Hawryshyn, C.W. 2004. Salmonid opsin sequences undergo positive selection and indicate an alternate evolutionary relationship in oncorhynchus. *J Mol Evol* **58**: 400–412.

Day, T.H., Hillier, P.C. & Clarke, B. 1974. Properties of genetically polymorphic isozymes of alcohol dehydrogenase in Drosophila melanogaster. *Biochemical genetics* **11**: 141-153.

Delport, W., Poon, A.F.Y., Frost, S.D.W. & Kosakovsky Pond, S.L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**: 2455–2457.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., *et al*. 2011. *Geneious v5.* Available: http://www.geneious.com.

Duffy, V.B. & Bartoshuk, L.M. 2000. Food Acceptance and Genetic Variation in Taste. *Journal of the American Dietetic Association* **100**: 647–655.

Ewens, W.J. & Gillespie, J.H. 1974. Some Simulation Results for the Neutral Allele Model, with Interpretations. *Theoretical Population Biology* **6**: 35-57.

Faurion, A., Saito, S. & Macleod, P. 1980. Sweet Taste Involves Several Distinct Receptor Mechanisms. *Chemical senses* **5**: 107–121.

Feiler, R., Bjornson, R., Kirschfeld, K., Mismer, D., Rubin, G.M., Smith, D.P., *et al.* 1992. Ectopic expression of ultraviolet-rhodopsins in the blue photoreceptor cells of Drosophila: visual physiology and photochemistry of transgenic animals. *J. Neurosci.* **12**: 3862–3868.

Flordellis, C., Manolis, A., Scheinin, M. & Paris, H. 2004. Clinical and pharmacological significance of α 2-adrenoceptor polymorphisms in cardiovascular diseases. *International Journal of Cardiology* **97**: 367–372.

Francuski, L., Milankov, V. & Ludoški, J. 2016. Genetic and phenotypic variation in central and northern European populations of Aedes (Aedimorphus) vexans (Meigen, 1830)(Diptera, Culicidae). *Journal of Vector Ecology* **41**: 160-171.

Frentiu, F.D., Bernard, G.D., Sison-Mangus, M.P., Brower, A.V.Z. & Briscoe, A.D. 2007. Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. *Mol. Biol. Evol.* **24**: 2016–2028.

Grant, B.S., Owen, D.F. & Clarke, C.A. 1996. Parallel rise and fall of melanic peppered moths in America and Britain. *Journal of Heredity* **87**: 351-357.

Graur, D. & Li, W.H. 2000. *Fundamentals of Molecular Evolution*. Sinauer.

Hayden, E.J., Ferrada, E. & Wagner, A. 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* **474**: 92–95.

Hoekstra, H.E., Hirschmann, R.J., Bundey, R.A., Insel, P.A. & Crossland, J.P. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101–104.

Hoffmann, M., Tripathi, N., Henz, S.R., Lindholm, A.K., Weigel, D., Breden, F., *et al.* 2007. Opsin gene duplication and diversification in the guppy, a model for sexual selection. *Proceedings of the Royal Society B: Biological Sciences* **274**: 33–42.

Huang, X.Q., Börner, A., Röder, A.S., Ganal, M.W. 2002. Assessing genetic diversity of wheat (Triticum aestivum L.) germplasm using microsatellite markers. *TAG Theoretical and Applied Genetics* **105**: 699–707.

Jerlov, N.G. 1976. *Jerlov: Marine Optics,* 229.

Johanson, U., West, J., Lister, C. & Michaels, S. 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science* **290**: 344-347.

Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**: 3059–3066.

Kim, K., Kotov, A.A. & Taylor, D.J. 2006. Hormonal induction of undescribed males resolves cryptic species of cladocerans. *Proceedings of the Royal Society B: Biological Sciences* **273**: 141–147.

Kim, U.K. & Drayna, D. 2004. Genetics of individual differences in bitter taste perception: lessons from the PTC gene. *Clinical Genetics* **67**: 275–280.

Kimura, M. & Ohta, T. 1973. The age of a neutral mutant persisting in a finite population. *Genetics* **75**: 199–212.

Kojima, D. 1997. A Novel Go-mediated Phototransduction Cascade in Scallop Visual Cells. *Journal of Biological Chemistry* **272**: 22979–22982.

Lell, B., May, J., Schmidt-Ott, R.J., Lehman, L.G., Luckner, D., Greve, B., *et al.* 1999. The role of red blood cell polymorphisms in resistance and susceptibility to malaria. *Clin. Infect. Dis.* **28**: 794–799.

Li, D., Su, Y., Tu, J., Wei, R., Fan, X. & Yin, H. 2016. Evolutionary conservation of the circadian gene timeout in Metazoa. *Animal Biology* **66**: 1-11

Li, Z., Gan, X. & He, S. 2009. Distinct Evolutionary Patterns Between Two Duplicated Color Vision Genes Within Cyprinid Fishes. *J Mol Evol*. **69**: 346-359.

Lynch, M. & Hill, W.G. 1986. Phenotypic evolution by neutral mutation. *Evolution* **40**: 915-935.

Lythgoe, J.N. 1968. Visual pigments and visual range underwater. *Vision Res.* **8**: 997–1011.

Matsunami, H., Montmayeur, J.P. & Buck, L.B. 2000. A family of candidate taste receptors in human and mouse. *Nature* **404**: 601–604.

Miyoshi, M.A., Abe, K. & Emori, Y. 2001. IP3 receptor type 3 and PLC β 2 are co-expressed with taste receptors T1R and T2R in rat taste bud cells. *Chemical senses*. **26**: 259-265.

Mock, K.E., Brim Box, J.C., Chong, J.P., Furnish, J. & Howard, J.K. 2013. Comparison of population genetic patterns in two widespread freshwater mussels with contrasting life histories in western North America. *Molecular Ecology* **22**: 6060–6073.

Muller, H.J., 1949. Reintegration of the symposium on genetics, paleontology, and evolution. *Genetics, paleontology and evolution*, pp.421-445.

Nasi, E. & Gomez, M.D.P. 2014. Melanopsin-mediated light-sensing in amphioxus. *Communicative & Integrative Biology* **2**: 441–443.

O'Connor, M., Garm, A., Marshall, J.N., Hart, N.S., Ekstrom, P., Skogh, C., *et al.* 2010. Visual pigment in the lens eyes of the box jellyfish Chiropsella bronzie. *Proceedings of the Royal Society B: Biological Sciences* **277**: 1843–1848.

Odierna, G., Aprea, G., Barucca, M., Canapa, A., Capriglione, T. & Olmo, E. 2006. Karyology of the Antarctic scallop Adamussium colbecki, with some comments on the karyological evolution of pectinids. *Genetica* **127**: 341–349.

Oprian, D.D., Molday, R.S. & Kaufman, R.J. 1987. Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proceedings of the National Academy of Sciences* **84**: 8874-8878.

Paaby, A.B. & Rockman, M.V. 2014. Cryptic genetic variation: evolution's hidden substrate. *Nat. Rev. Genet.* **15**: 247–258.

Parry, C.M., Erkner, A. & le Coutre, J. 2004. Divergence of T2R chemosensory receptor families in humans, bonobos, and chimpanzees. *Proceedings of the National Academy of Sciences* **101**: 14830–14834.

Pointer, M.A. & Mundy, N.I. 2008. Testing whether macroevolution follows microevolution: are colour differences among swans (Cygnus) attributable to variation at the MCIR locus? *BMC Evol. Biol.* **8**: 249.

Pond, S.L.K. & Frost, S.D.W. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**: 2531–2533.

Porter, M.L., Cronin, T.W., McClellan, D.A. & Crandall, K.A. 2006. Molecular Characterization of Crustacean Visual Pigments and the Evolution of Pancrustacean Opsins. *Mol. Biol. Evol.* **24**: 253–268.

Puslednik, L. & Serb, J.M. 2008. Molecular phylogenetics of the Pectinidae (Mollusca: Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology. *Mol. Phylogenet. Evol.* **48**: 1178–1188.

Reed, D.R., Tanaka, T. & McDaniel, A.H. 2006. Diverse tastes: Genetics of sweet and bitter perception. *Physiology & Behavior* **88**: 215–226.

Rohner, N., Jarosz, D.F., Kowalko, J.E., Yoshizawa, M., Jeffery, W.R., Borowsky, R.L., *et al.* 2013. Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish. *Science* **342**: 1372–1375.

Rosenblum, E.B., Römpler, H., Schöneberg, T. & Hoekstra, H.E. 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proc. Natl. Acad. Sci. U.S.A.* **107**: 2113–2117.

Sakmar, T.P., Menon, S.T., Marin, E.P. & Awad, E.S. 2002. Rhodopsin: insights from recent structural studies. *Annu. Rev. Biophys. Biomol. Struct.* **31**: 443–484.

Scheffler, K., Martin, D.P. & Seoighe, C. 2006. obust inference of positive selection from recombining coding sequences. *Bioinformatics* **22**: 2493–2499.

Schlichting, C.D. 2008. Hidden Reaction Norms, Cryptic Genetic Variation, and Evolvability. *Annals of the New York Academy of Sciences* **1133**: 187–203.

Serb, J.M., Porath-Krause, A.J. & Pairett, A.N. 2013. Uncovering a Gene Duplication of the Photoreceptive Protein, Opsin, in Scallops (Bivalvia: Pectinidae). *Integrative and Comparative Biology*, doi: 10.1093/icb/ict063.

Shi, L., Liu, X., Liu, B., Zhao, X., Wang, L., Li, J., *et al.* 2009. Identifying neutral allele Sb at pollen-sterility loci in cultivated rice with Oryza rufipogon origin. *Chin. Sci. Bull.* **54**: 3813–3821.

Smadja, C., Shi, P., Butlin, R.K. & Robertson, H.M. 2009. Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, Acyrthosiphon pisum. *Mol. Biol. Evol.* **26**: 2073–2086.

Spady, T.C. 2005. Adaptive Molecular Evolution in the Opsin Genes of Rapidly Speciating Cichlid Species. *Mol. Biol. Evol.* **22**: 1412–1422.

Speiser, D.I. & Johnsen, S. 2008. Comparative morphology of the concave mirror eyes of scallops (Pectinoidea). *Amer. Malac. Bull*. **26**: 27–33.

Speiser, D.I., Loew, E.R. & Johnsen, S. 2011. Spectral sensitivity of the concave mirror eyes of scallops: potential influences of habitat, self-screening and longitudinal chromatic aberration. *J. Exp. Biol.* **214**: 422–431.

Storz, J.F., Sabatino, S.J., Hoffmann, F.G., Gering, E.J., Moriyama, H., Ferrand, N., *et al.* 2007. The molecular basis of high-altitude adaptation in deer mice. *PLoS Genet.* **3**: e45.

Tabarini, C.L. 1984. Induced triploidy in the bay scallop, Argopecten irradians, and its effects on growth and gametogenesis. *Aquaculture*. **42**: 151-160.

Takahashi, Y. & Ebrey, T.G. 2003. Molecular basis of spectral tuning in the newt short wavelength sensitive visual pigment. *Biochemistry* **42**: 6025–6034.

Terakita, A., Hariyama, T., Tsukahara, Y., Katsukura, Y. & Tashiro, H. 1993. Interaction of GTP-binding protein Gq with photoactivated rhodopsin in the photoreceptor membranes of crayfish. *FEBS Lett.* **330**: 197–200.

Terakita, A., Tsukamoto, H. & Koyanagi, M. 2008. Expression and comparative characterization of Gq-coupled invertebrate visual pigments and melanopsin. *Journal of Neurochemistry* **105**: 883-890.

Todesco, M., Balasubramanian, S., Hu, T.T., Traw, M.B., Horton, M., Epple, P., *et al.* 2010. Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. *Nature* **465**: 632–636.

Townson, S.M., Chang, B.S.W., Salcedo, E., Chadwell, L.V., Pierce, N.E. & Britt, S.G. 1998. Honeybee Blue- and Ultraviolet-Sensitive Opsins: Cloning, Heterologous Expression in Drosophila, and Physiological Characterization. *The Journal of Neuroscience* **18**: 2412-2422.

Webster, T.P., Selander, R.K. & Yang, S.Y. 1972. Genetic Variability and Similarity in the Anolis Lizards of Bimini. *Evolution* **26**: 523-535.

West-Eberhard, M.J. 1983. Sexual selection, social competition, and speciation. *Quarterly Review of Biology* **58**: 155-183.

Wilkie, S.E., Robinson, P.R., Cronin, T.W., Poopalasundaram, S., Bowmaker, J.K. & HUNT, D.M. 2000. Spectral Tuning of Avian Violet- and Ultraviolet-Sensitive Visual Pigments. *Biochemistry* **39**: 7895–7901.

Yang, S.Y. & Patton, J.L. 1981. Genic Variability and Differentiation in the Galapagos Finches. *The Auk* **98**: 230-242.

Yokoyama, S. 2002. Molecular evolution of color vision in vertebrates. *Gene* **300**: 69–78.

Yokoyama, S. & Radlwimmer, F.B. 1999. The molecular genetics of red and green color vision in mammals. *Genetics* **153**: 919–932.

Yokoyama, S., Tada, T., Zhang, H. & Britt, L. 2008. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **105**: 13480–13485.

Table 4-1 Species, number of different nucleotide and amino acid sequences, and GenBank accession numbers of nucleotide sequences

Species	*opnGq1* nt seq	OPNGq1 aa seq	*opnGq2* nt seq	OPNGq2 aa seq	*opnGq1*	*opnGq2*
Ctenoides annulatus	4	4	2	2	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	
					NCBI #	
Bractechlamys antillarum	5	4	1	1	NCBI #	NCBI #
					NCBI #	
					NCBI #	
					NCBI #	
					NCBI #	
Ylistrum balloti	1	1	2	1	NCBI #	NCBI #
						NCBI #
Leptopecten bavayi	4	4	6	4	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
						NCBI #
Spathochlamys benedicti	4	4	3	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	
Mimachlamys cloacata	2	2	1	1	NCBI #	NCBI #
					NCBI #	
Adamussium colbecki	3	3	4	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
Propeamussium dalli	1	1	2	2	NCBI #	NCBI #
						NCBI #
Veprichlamys empressae	3	3	4	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
Pecten fumatus	2	2	4	4	NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
						NCBI #
Crassodoma gigantea	1	1	–	–	NCBI #	–
Chlamys hastata	4	2	3	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	
Spondylus ictericus	3	3	2	2	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	
Argopecten irradians	1	1	2	2	NCBI #	NCBI #
						NCBI #
Ylistrum japonicum	5	4	4	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	

Table 4-1 (continued)

Species	*opnGq1* nt seq	OPNGq1 aa seq	*opnGq2* nt seq	OPNGq2 aa seq	*opnGq1*	*opnGq2*
Laevichlamys lemniscata	4	3	4	2	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
Placopecten magellanicus	2	2	4	4	NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
						NCBI #
Semipallium marybellae	5	4	1	1	NCBI #	NCBI #
					NCBI #	
					NCBI #	
					NCBI #	
					NCBI #	
Pecten maximus	2	2	1	1	NCBI #	NCBI #
					NCBI #	
Talochlamys multistriata	6	6	3	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	
					NCBI #	
					NCBI #	
Aequipecten opercularis	4	3	4	4	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
Caribachlamys ornata	5	5	4	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	
Amusium papyraceum	3	3	1	1	NCBI #	NCBI #
					NCBI #	
					NCBI #	
Amusium pleuronectes	4	3	1	1	NCBI #	NCBI #
					NCBI #	
					NCBI #	
					NCBI #	
Decatopecten plica	3	2	1	1	NCBI #	NCBI #
					NCBI #	
					NCBI #	
Talochlamys pusio	2	2	4	4	NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
						NCBI #
Chlamys rubida	3	2	5	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
						NCBI #
Mimachlamys sanguinea	4	3	3	1	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	
Pseudamussium septemradiatum	4	3	3	3	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	

Table 4-1 (continued)

Species	*opnGq1* nt seq	OPNGq1 aa seq	*opnGq2* nt seq	OPNGq2 aa seq	*opnGq1*	*opnGq2*
Pedum spondyloideum	2	2	3	3	NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
Cryptopecten vesiculosus	4	4	5	4	NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
Euvola vogdesi	1	1	4	4	NCBI #	NCBI #
						NCBI #
						NCBI #
						NCBI #
Mizuhopecten yessoensis	2	2	5	4	NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #
						NCBI #
						NCBI #
Euvola ziczac	2	2	3	3	NCBI #	NCBI #
					NCBI #	NCBI #
						NCBI #

Table 4-2 Single nucleotide polymorphisms and percent pairwise identity of sequences

Species	***opnGq1***			***opnGq2***		
	Nucleotide Alleles	Positions with SNPs	Percent Pairwise Identity	Nucleotide Alleles	Positions with SNPs	Percent Pairwise Identity
Ctenoides annulatus	4	6	>99	2	80	~84
Bractechlamys antillarum	5	9	>99	1	_	_
Ylistrum balloti	1	_	_	2	1	>99
Leptopecten bavayi	4	61	89 – 97	6	59	~89 - >99
Spathochlamys benedicti	4	7	99 - >99	3	53	~90 - >99
Mimachlamys cloacata	2	3	>99	1	_	_
Adamussium colbecki	3	10	98 – 99	4	6	99 - >99
Propeamussium dalli	1	_	_	2	75	83.7
Veprichlamys empressae	3	3	>99	4	64	~85 - >99
Pecten fumatus	2	1	>99	4	76	~85 - >99
Crassodoma gigantea	1	_	_	NA	NA	NA
Chlamys hastata	4	5	>99	3	6	99 - >99
Spondylus ictericus	3	3	>99	2	9	98.3
Argopecten irradians	1	_	_	2	2	>99
Ylistrum japonicum	5	13	>99	4	5	98.9 - >99
Laevichlamys lemniscata	4	7	~99 - >99	4	6	>99
Placopecten magellanicus	2	5	>99	4	73	85.1 - >99
Semipallium marybellae	5	9	~99 - >99	1	_	_
Pecten maximus	2	2	>99	1	_	_
Talochlamys multistriata	6	15	~98 - >99	3	106	84 – 85.1
Aequipecten opercularis	4	4	>99	4	116	84.5 – 86.9
Caribachlamys ornata	5	70	86 - >99	4	16	97.1 - >99
Amusium papyraceum	3	5	99 - >99	1	_	_
Amusium pleuronectes	4	3	>99	1	_	_
Decatopecten plica	3	11	~98 - >99	1	_	_
Talochlamys pusio	2	11	~98	4	3	>99
Chlamys rubida	3	3	>99	5	59	88.4 - >99
Mimachlamys sanguinea	4	3	>99	3	2	>99
Pseudamussium septemradiatum	4	4	>99	3	3	>99
Pedum spondyloideum	2	5	99	3	97	84.1 – 85.7
Cryptopecten vesiculosus	4	6	99 - >99	5	5	>99
Euvola vogdesi	1	_	_	4	8	98.9 - >99
Mizuhopecten yessoensis	2	8	~98	5	6	99 - >99
Euvola ziczac	2	9	~98	3	3	>99

Table 4-3 dN/dS ratios from randomized samplings of available nucleotide sequences

	Randomization	Ln-likelihood	Parameters	dN/dS ratio
opnGq1	Run 1	-2444.43	59	0.10
	Run 2	-2301.02	57	0.06
	Run 3	-2459.01	57	0.09
opnGq2	Run 1	-3797.49	71	0.06
	Run 2	-3721.24	69	0.06
	Run 3	-3626.35	71	0.07

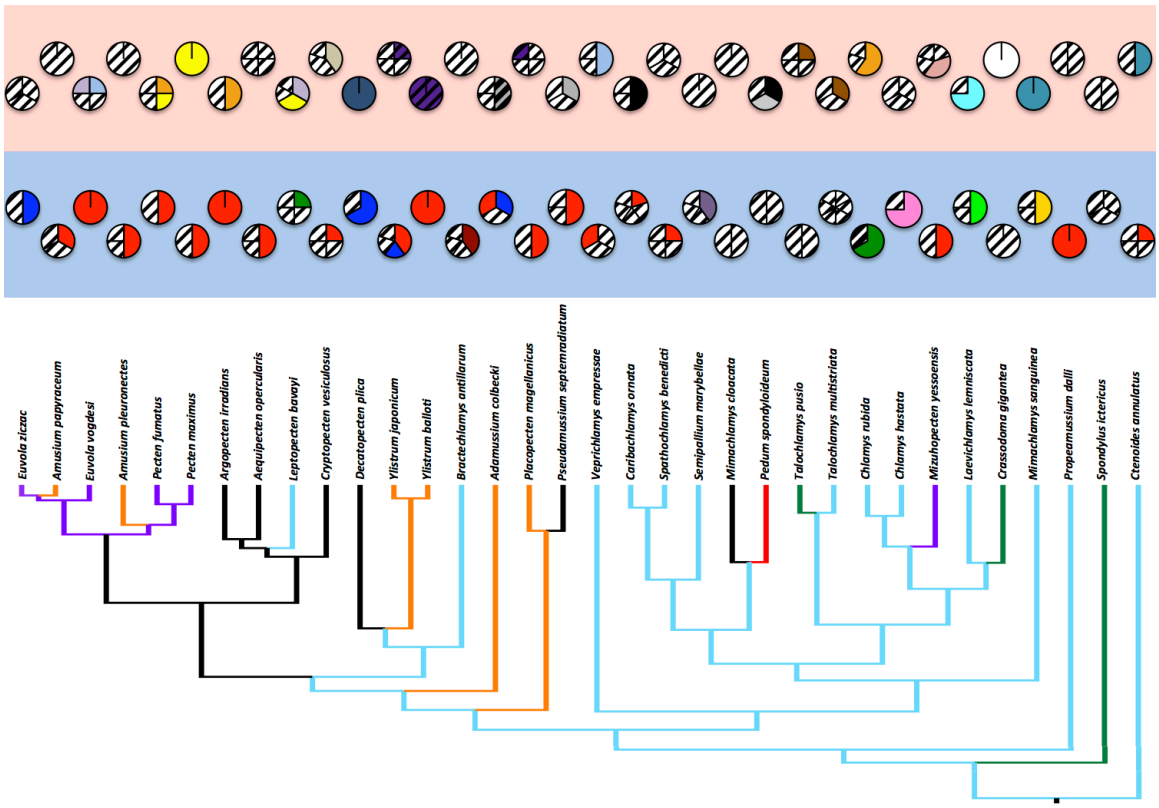


Figure 4-1 OPNGq1 and OPNGq2 amino acid sequence variation across scallops

Pie charts represent the total number of different amino acid sequences translated from nucleotide sequences for each corresponding species. Each color represents a different amino acid sequence. Blue shading behind pie charts represents sequences belonging to OPNGq1 while red shading behind pie charts represents sequences belonging to OPNGq2. Black and white striped coloration represents a single, unique amino acid sequence that is not shared with any other sequences within or between species. Solid white pie chart represents an absence of any available sequence. Colors along the branches of the phylogeny represent the six life habit classes (specified in Alejandrino *et al*., 2011) exhibited by scallops; orange = gliding, purple = recessing, blue = byssal attaching, black = free-living, green = cementing, red = nestling.

CHAPTER V

CONCLUSIONS

In this dissertation, I reveal that scallop G_q-opsins have increased genetic variation through gene duplication. As a result, photosensory system expansion could lead to functional diversity and light-sensing capabilities in scallops. Within Chapter II, I identified four *Air-opsGq* paralogs that appear capable to form viable photopigments in the scallop *Argopecten irradians*. Differential levels of gene expression indicate an even finer spatial partitioning of *Air-opsGqs* could occur across ocular and extra-ocular structures in the adult, suggesting there have been changes in the regulatory regions of scallop G_q-opsin paralogs. Spatial patterning and expression level differences among the G_q-opsin paralogs suggest they have neofunctionalized since duplication and are used in different biological contexts. *Air-opnGq3* and *Air-opnGq4* may preferentially be employed in eyes, while others *Air-opnGq1* and *Air-opnGq2* are used for both ocular and extra-ocular based functions.

Using a species-level gene comparison, I demonstrated two G_q-opsin paralogs are under different evolutionary trajectories across the Pectinoida superfamily in Chapter III. *opnGq2* was identified to have twice the rate of purifying selection as *opnGq1*, which could be explained by selection for the sequence variant B from the scallops that use the gliding behavior. Eight positions were identified as potential candidates to further explain the difference in the evolutionary rates at the molecular level. One residue change, D288N, is of particular interest because all six gliding species maintain a radical difference at that position compared to all other species. These data suggest that a

visually-mediated behavior is molding the molecular composition of a G_q-opsin paralog, found in the scallop.

The fourth chapter, I revealed an unprecedented amount of genetic variation in the form of *opnGq1* and *opnGq2* alleles from across scallops. Using an unedited sample set, I demonstrated how removing alleles could bias tests of molecular evolution, such as dN/dS ratios, which could impact additional studies (e.g., strength of selection or phylogenetic relationship). Once translated, the alleles produced multiple nonsynonymous amino acid sequences, suggesting this may be hidden or cryptic variation within a species that could be advantageous to a population as it can facilitate an opportunity for adaptation to novel environmental conditions. Coupled with the data collected in Chapters II and III, I uncovered novel genetic and functional diversity in the light-sensing structures of the scallop. These data provide a basis to study how gene duplication may coincide with the evolution of complex organs, specifically the eye, by exploring the different ways neofunctionalization of G_q-opsins may occur, by teasing apart the environmental forces acting on the molecular components, and by incorporating cryptic variation in ever-changing environments and climates.

My work demonstrates the complex nature of opsins in molluscs, but this system is ripe for more studies looking at the visual evolution of molluscs to further their impact on the fields of molecular, sensory, and evolutionary biology. Future tests, such as site-directed mutagenesis and expression *in vitro*, that include amino acid positions which are not frequently considered important in studies of invertebrate spectral tuning, may be critical to reveal important patterns of molecular evolution in G_q-opsins and their role in spectral tuning. There is great need for increased taxonomic sampling of intraspecific

opsin diversity in non-arthropod invertebrates to understand the plasticity of G_q-opsins across animals which may further address questions of duplication and divergence that led to the evolution of visual systems.